Docket No.: GNE.3230R1C85

December 20, 2005

Please Direct All Correspondence to Customer Number **30313**

TRANSMITTAL LETTER

APPEAL BRIEF

Applicant	:	Goddard, et al. (as amended)
App. No	:	10/063,617
Filed	:	May 3, 2002
For	:	SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
Examiner	:	Romeo, David S.
Art Unit	:	1647

CERTIFICATE OF MAILING

I hereby certify that this correspondence and all marked attachments are being deposited with the United States Postal Service as first-class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on

December 20, 2005
(Date)

AnneMarie Kaiser, Reg. No. 37,649

Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Transmitted herewith for filing in the above-identified application are the following enclosures:

(X) Appeal Brief in 83 pages.

FILING FEES:

FEE CALCULATION				
FEE TYPE		**FEE CODE**	**CALCULATION**	**TOTAL**
Appeal Brief	*41.20(b)(2)*	1402 ($500)		$500
1 Month Extension	*1.17(a)(1)*	1251 ($120)		$
2 Month Extension	*1.17(a)(2)*	1252 ($450)		$
3 Month Extension	*1.17(a)(3)*	1253 ($1,020)		$
			TOTAL FEE DUE	$500

(X) A copy of evidence cited in Appellant's Brief and listed in Appendix B.

(X) A check in the amount of $500 is enclosed.

(X) Return prepaid postcard.

12/28/2005 HLE333 00000026 10063617

01 FC:1402 500.00 OP

Docket No.	:	GNE.3230R1C85	**Customer No.: 30,313**
Application No.	:	10/063,617	
Filing Date	:	May 3, 2002	

Please charge any additional fees, including any fees for additional extension of time, or credit overpayment to Deposit Account No. 11-1410.

Dated: December 20, 2005

AnneMarie Kaiser
Registration No. 37,649
Attorney of Record
Customer No. 30,313
(619) 235-8550

2226855
122005

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	:	Goddard, et al.
Appl. No.	:	10/063,617
Filed	:	May 3, 2002
For	:	SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
Examiner	:	Romeo, David S.
Group Art Unit	:	1647

CERTIFICATE OF MAILING

I hereby certify that this correspondence and all marked attachments are being deposited with the United States Postal Service as first-class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on

December 20, 2005
(Date)

AnneMarie Kaiser, Reg. No. 37,649

ON APPEAL TO THE BOARD OF PATENT APPEALS AND INTERFERENCES
APPELLANT'S BRIEF

Mail Stop Appeal Brief – Patents
COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

The Applicants appeal the rejection of Claims 6-8 and 11-17 in the above-captioned patent application. These claims were rejected in a final Office Action dated July 26, 2005. In response to Applicants' Amendment After Final, the Examiner issued an Advisory Action dated October 12, 2005. Applicants mailed a Notice of Appeal October 25, 2005.

I. REAL PARTY IN INTEREST

Pursuant to 37 C.F.R. 41.37(c)(1), Appellants hereby notify the Board of Patent Appeals and Interferences that the real party in interest is the assignee of record for this application, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080.

II. RELATED APPEALS AND INTERFERENCES

A Notice of Appeal has been filed in the related Application Nos. 10/063,570, 10/063,591, 10/063,607, 10/063,530, 10/063,617 and 10/063,640. A Notice of Appeal and an Appeal Brief have also been filed in the related Application Nos. 10/063,519; 10/063,524; 10/063,534; 10/063,540; 10/063,578; 10/063,584; 10/063,586; 10/063,587; 10/063,592; 10/063,616; 10/063,648; 10/063,652; 10/063,653; 10/063,659; 10/063,660; and 10/063,661. Appellants are unaware of any other related appeals or interferences.

III. STATUS OF THE CLAIMS

The above-captioned application was filed with Claims 1-13. Appellants cancelled Claims 1-3 and 9-10 and added new Claims 14-17 in an Amendment and Response to Office Action dated April 29, 2005. The Examiner rejected Claims 4-8 and 11-17 in a Final Office Action dated July 26, 2005. Appellants filed an Amendment After Final Office Action on September 23, 2005, amending Claim 12 to depend from Claim 6, and canceling Claims 4 and 5. The Examiner issued an Advisory Action dated October 12, 2005, which indicated that the amendments in the Amendment After Final Office Action were entered for purposes of appeal. Accordingly, Claims 6-8 and 11-17 are the subject of this appeal. The claims attached hereto as Appendix A reflect the claims as amended by the Amendment After Final Office Action.

IV. STATUS OF AMENDMENTS

Appellants mailed an Amendment After Final Office Action on September 23, 2005, amending Claim 12 to depend from Claim 6, and canceling Claims 4 and 5. The Examiner issued an Advisory Action dated October 12, 2005, which indicated that the amendments in the Amendment After Final Office Action were entered for purposes of appeal.

V. SUMMARY OF THE CLAIMED SUBJECT MATTER

The claimed subject matter relates to isolated polypeptides related to the polypeptide having SEQ ID NO:110. Independent Claims 6 and 14 read:

6. An isolated polypeptide comprising:

(a) the amino acid sequence of the polypeptide of SEQ ID NO: 110;

(b) the amino acid sequence of the polypeptide of SEQ ID NO: 110, lacking its associated signal peptide; or

(c) the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203535.

14. An isolated polypeptide having at least 95% amino acid sequence identity to:

(a) the amino acid sequence of the polypeptide of SEQ ID NO: 110;

(b) the amino acid sequence of the polypeptide of SEQ ID NO: 110, lacking its associated signal peptide; or

(c) the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203535;

wherein said isolated polypeptide or a fragment thereof can be used to generate an antibody which can be used to specifically detect the polypeptide of SEQ ID NO: 110 in esophageal tissue samples.

Various aspects of the claimed polypeptides are described in the specification at, for example, paragraphs [0012]-[0017], [0021], [0023], [0026], [0135]-[0136], [0198]-[0208], [0221], [0225], [0229], [0231], [0247], [0253]-[0271], [0280]-[0282], [0336], [0364], [0367], [0369], [0454]-[0499] and [0530], and Figure 110. SEQ ID NO:110 is disclosed in the Sequence Listing appended to the application.

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL AND GROUPING OF CLAIMS

A. Grounds of Rejection on Appeal

The Examiner has rejected pending Claims 6-8 and 11-17 under 35 U.S.C. §101, stating that the claimed invention is not supported by either a specific and substantial asserted utility or a well-established utility. *Final Office Action* at 2.

The Examiner has rejected pending Claims 6-8 and 11-17 under 35 U.S.C. §112, first paragraph as lacking an enabling disclosure, "since the claimed invention is not supported by

either a specific and substantial utility or a well established utility." *Id.* at 24. In addition, pending claims 12-17 are rejected as failing to comply with the enablement requirement because "the scope of enablement provided to the skilled artisan by the disclosure is not commensurate with the scope of the protection sought by the claims." *Id.* at 25-27.

Finally, the Examiner has rejected pending Claims 12-17 under 35 U.S.C. § 112, first paragraph, as lacking an adequate written description, stating that the rejected claims contain "subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor, at the time the application was filed, had possession of the claimed invention." *Id.* at 27.

B. Grouping of Claims

1. *Utility Rejection – Claims 6-8 and 11-17*

For purposes of the utility rejection under 35 U.S.C. § 101, Claims 6-8 and 11-17 can be considered as a group.

2. *Enablement Rejection – Claims 6-8 and 11-17*

a. Group 1 – Claims 6-8 and 11-13

For purposes of the enablement rejection under 35 U.S.C. § 112, first paragraph, Claims 6-8 and 11-13 can be considered as a group.

b. Group 2 – Claims 14 and 16-17

For purposes of the enablement rejection under 35 U.S.C. § 112, first paragraph, Claims 14 and 16-17 can be considered as a group.

c. Group 3 – Claim 15

For purposes of the enablement rejection under 35 U.S.C. § 112, first paragraph, Claim 15 can be considered individually.

3. ***Written Description Rejection – Claims 12-17***

a. ***Group 1 – Claims 12-13***

For purposes of the written description rejection under 35 U.S.C. § 112, first paragraph, Claims 12-13 can be considered as a group.

b. ***Group 2 – Claim 14 and 16-17***

For purposes of the written description rejection under 35 U.S.C. § 112, first paragraph, Claims 14 and 16-17 can be considered as a group.

c. ***Group 3 – Claim 15***

For purposes of the written description rejection under 35 U.S.C. § 112, first paragraph, Claim 15 can be considered individually.

VII. APPELLANTS' ARGUMENT

A. Summary of Arguments

1. ***Utility Rejection***

The first issue before the Board is whether Appellants have asserted at least one "specific, substantial, and credible utility" for the claimed subject matter. *See* Examination Guidelines, 66 Fed. Reg. 1092 (2001). Appellants have asserted that the claimed polypeptides related to the polypeptide of SEQ ID NO:110 (the PRO1753 polypeptide) are useful as diagnostic tools for cancer, particularly for esophageal cancer. This asserted utility is specific, substantial, and credible.

Briefly stated, Appellants' asserted utility is based on the disclosure in Example 18 of the instant application that the mRNA encoding the PRO1753 polypeptide is more highly expressed in esophageal tumor tissue compared to normal esophageal tissue. It is well-established that there is a reasonable correlation between changes in mRNA level for a particular gene and a corresponding change in the level of expression of the encoded polypeptide, such that increasing or decreasing the amount of mRNA for a particular gene leads to a corresponding increase or decrease in the amount of the encoded protein. Thus, one of skill in the art would be more likely

than not to believe that, like the PRO1753 mRNA, the PRO1753 protein is more highly expressed in esophageal tumor tissue compared to normal esophageal tissue. This differential expression of PRO1753 polypeptide is useful for distinguishing esophageal tumor tissue from its normal tissue counterpart. Therefore, the claimed polypeptides related to the PRO1753 polypeptide have a specific, substantial and credible utility as diagnostic tools for cancer, particularly esophageal cancer, as is explained in more detail below.

2. ***Enablement Rejection***

The second issue before the Board is whether Appellants have enabled the pending claims such that one of skill in the art would be able to make and use the claimed invention. The Examiner has rejected pending Claims 6-8 and 11-17 under 35 U.S.C. §112, first paragraph as lacking an enabling disclosure, because the claimed invention is not supported by either a specific and substantial utility or a well established utility. *Id.* at 24. In addition, pending claims 12-17 are rejected because "the scope of enablement provided to the skilled artisan by the disclosure is not commensurate with the scope of the protection sought by the claims." *Id.* at 25-27.

Appellants submit that Claims 6-8 and 11-17 are enabled such that one of skill in the art could make and use the claimed polypeptides without undue experimentation. First, for the reasons discussed in detail below, the claimed invention is supported by a specific, substantial and credible utility. Accordingly, Appellants maintain that the specification enables one skilled in the art to make and use the claimed polypeptides.

With respect to Claims 12-13, how to make the polypeptide of SEQ ID NO:110 and the polypeptide encoded by the cDNA deposited under ATCC accession number 203535 is within the skill in the art. Similarly, with respect to Claims 14-17, it is well within the skill of those in the art to make polypeptides that are at least 95% identical to SEQ ID NO:110 and the polypeptide encoded by ATCC 203535, and it is well within the knowledge of those skilled in the art how to make antibodies which are specific to a disclosed sequence. *See In re Wands*, 858 F.2d 731 (reversing the Board's decision of non-enablement and holding that as of 1980, undue

experimentation was not required to make high-affinity monoclonal antibodies to a target peptide).

Appellants assert that the claimed polypeptides are useful as diagnostic tools for cancer, particularly esophageal cancer. This use is based on the disclosure in Example 18 of the instant application that the nucleic acid encoding the PRO1753 polypeptide is overexpressed at least two-fold in esophageal tumor compared to normal esophageal tissue. As detailed below, it is well-established that changes in the expression level of mRNA is correlated with a change in the expression level of the encoded polypeptide, and thus it is likely that the PRO1753 polypeptide is overexpressed in esophageal tumors. Thus, based on the disclosure in the application, one of skill in the art would be able to use the claimed polypeptides as diagnostic tools to distinguish suspected esophageal tumors from normal esophageal tissue without undue experimentation.

3. ***Written Description Rejection***

The third issue before the Board is whether the claimed subject matter is described in the specification in such a way as to reasonably convey to one skilled in the art that the inventors had possession of the claimed invention at the time the application was filed. The Examiner has rejected pending Claims 12-17 under 35 U.S.C. §112, first paragraph, as lacking an adequate written description. *Id.* at 27.

The Examiner rejects claims 12-13 based on their dependency from Claim 4, arguing that the limitation "wherein said isolated polypeptide is more highly expressed..." is in adequate because "no information is provided in the differential analysis of PRO1753 polynucleotide expression regarding the level of expression, activity, or role in cancer of the PRO1753 polypeptide." *Id.* at 27. In rejecting Claims 14-17, the Examiner argues that "because the specification does not describe any biological activity of the claimed polypeptides and because the claims are not limited to any specific biological activity of the claimed polypeptides, the present claims are not analogous to example 14 of the written description guidelines." *Id.* at 28-29.

Appellants submit that they have satisfied the written description requirement for rejected Claims 12-17. The Examiner's argument regarding Claims 12-13 is moot in light of the

amendment of Claims 12-13 to depend from Claim 6, which the Examiner has not rejected as lacking an adequate written description.

As for the rejection of Claims 14-17, Applicants submit that the claims are adequately described based on the actual reduction to practice of SEQ ID NO:110, by specifying a high level of amino acid sequence identity, and by describing how to make and use antibodies to the disclosed sequence. Contrary to the Examiner's assertion, these facts are directly analogous to those of Example 14 of the Written Description Guidelines published by the PTO, since the claims are limited to polypeptides with the biological function of being "used to generate an antibody which can be used to specifically detect the polypeptide of SEQ ID NO: 110 in esophageal tissue samples." In addition, like the case *In re Wallach* where a single polypeptide sequence was sufficient written description for the vast number of nucleic acids which encode the polypeptide, the description of the single species SEQ ID NO:110 is sufficient to place the Appellants' in possession of the claimed genus because those of skill in the art recognize the correlation between polypeptide structure and the ability to generate specific antibodies. For these reasons as detailed below, Appellants submit that Claims 12-17 are adequately described such that one of skill in the art would recognize that the inventors had possession of the claimed invention at the time the application was filed.

B. Utility Rejection – Detailed Arguments

The first issue before the Board is whether Appellants have asserted at least one "specific, substantial, and credible utility." *See Examination Guidelines*, 66 Fed. Reg. 1092 (2001). Appellants have asserted that the claimed polypeptides related to SEQ ID NO:110 (the PRO1753 polypeptide) are useful as diagnostic tools for cancer, particularly for esophageal cancer. This asserted utility is specific, substantial, and credible, as is explained in more detail below.

1. *Utility – Legal Standard*

A "specific utility" is defined as utility which is "specific to the subject matter claimed," in contrast to "a general utility that would be applicable to the broad class of the invention." *See*

M.P.E.P. § 2107.01 I. For example, it is generally not enough to state that a nucleic acid is useful as a diagnostic tool without also identifying the condition that is to be diagnosed.

The requirement of "substantial utility" defines a "real world" use, and derives from the Supreme Court's holding in *Brenner v. Manson*, 383 U.S. 519, 534 (1966) stating that "[t]he basic *quid pro quo* contemplated by the Constitution and the Congress for granting a patent monopoly is the benefit derived by the public from an invention with substantial utility." In explaining the "substantial utility" standard, M.P.E.P. § 2107.01 cautions, however, that Office personnel must be careful not to interpret the phrase "immediate benefit to the public" or similar formulations used in certain court decisions to mean that products or services based on the claimed invention must be "currently available" to the public in order to satisfy the utility requirement. "Rather, any reasonable use that an applicant has identified for the invention that can be viewed as providing a public benefit should be accepted as sufficient, at least with regard to defining a 'substantial' utility." *M.P.E.P.* § 2107.01 (emphasis added).

Indeed, the Guidelines for Examination of Applications for Compliance With the Utility Requirement, set forth in M.P.E.P. § 2107 II(B)(1) gives the following instruction to patent examiners: "If the applicant has asserted that the claimed invention is useful for any particular practical purpose ... and the assertion would be considered credible by a person of ordinary skill in the art, do not impose a rejection based on lack of utility."

Finally, in assessing the credibility of the asserted utility, the M.P.E.P. states that "to overcome the presumption of truth that an assertion of utility by the applicant enjoys" the PTO must establish that it is "more likely than not that one of ordinary skill in the art would doubt (i.e., 'question') the truth of the statement of utility." *M.P.E.P.* § 2107.02 III A.

2. ***Utility – Burden of Proof***

It is well established that a specification which contains a disclosure of utility which corresponds in scope to the subject matter sought to be patented "must be taken as sufficient to satisfy the utility requirement of § 101 for the entire claimed subject matter unless there is reason for one skilled in the art to question the objective truth of the statement of utility or its scope." *In re Langer*, 503 F.2d 1380, 1391, 183 U.S.P.Q. 288, 297 (C.C.P.A. 1974). Thus "the PTO has the

initial burden of challenging a presumptively correct assertion of utility in the disclosure." *In re Brana*, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995). Only after the PTO provides evidence showing that one of ordinary skill in the art would reasonably doubt the asserted utility does the burden shift to the applicant to provide rebuttal evidence sufficient to convince such a person of the invention's asserted utility. *Id.*

3. ***Utility – Standard of Proof***

Compliance with 35 U.S.C. § 101 is a question of fact. *Raytheon v. Roper*, 724 F.2d 951, 956, 220 U.S.P.Q. 592, 596 (Fed. Cir. 1983). The evidentiary standard to be used throughout *ex parte* examination in setting forth a rejection is a preponderance of the evidence, or "more likely than not" standard. *In re Oetiker*, 977 F.2d 1443, 1445, 24 U.S.P.Q.2d 1443, 1444 (Fed. Cir. 1992). This is stated explicitly in the M.P.E.P.:

> [T]he applicant does not have to provide evidence sufficient to establish that an asserted utility is true "beyond a reasonable doubt." Nor must the applicant provide evidence such that it establishes an asserted utility as a matter of statistical certainty. Instead, evidence will be sufficient if, considered as a whole, it leads a person of ordinary skill in the art to conclude that the asserted utility is more likely than not true. *M.P.E.P.* § 2107.02, part VII (emphasis in original, citations omitted).

The Court of Appeals for the Federal Circuit has stated that the standard for satisfying the utility requirement is a low one:

> The threshold of utility is not high: An invention is "useful" under section 101 if it is capable of providing some identifiable benefit. *See Brenner v. Manson,* 383 U.S. 519, 534, 86 S.Ct. 1033, 16 L.Ed.2d 69 (1966); *Brooktree Corp. v. Advanced Micro Devices, Inc.*, 977 F.2d 1555, 1571 (Fed. Cir. 1992) ("To violate § 101 the claimed device must be totally incapable of achieving a useful result"); *Fuller v. Berger,* 120 F. 274, 275 (7th Cir. 1903) (test for utility is whether invention "is incapable of serving any beneficial end"). *Juicy Whip, Inc. v. Orange Bang, Inc.*, 185 F.3d 1364, 1366, 51 U.S.P.Q. 2d 1700 (Fed. Cir. 1999) (emphasis added).

The low threshold for satisfying the utility requirement is reflected in the standard set by the Federal Circuit for invalidating a patent based on a lack of utility: "[T]he fact that an invention has only limited utility and is only operable in certain applications is not grounds for finding lack of utility. Some degree of utility is sufficient for patentability. Further, the defense of non-utility

cannot be sustained without proof of total incapacity." *Envirotech Corp. v. Al George, Inc.*, 730 F.2d 753, 762, 221 U.S.P.Q. 473 (Fed. Cir. 1984) (emphasis added, citations omitted).

Because the standard for satisfying the utility requirement is so low, requiring total incapacity for a finding of no utility, the M.P.E.P. cautions that:

> Rejections under 35 U.S.C. 101 have been *rarely* sustained by federal courts. Generally speaking, in these *rare* cases, the 35 U.S.C. 101 rejection was sustained [] because the applicant ... asserted a utility that could only be true if it violated a scientific principle, such as the second law of thermodynamics, or a law of nature, or was wholly inconsistent with contemporary knowledge in the art. *M.P.E.P.* § 2107.02 III B., citing *In re Gazave*, 379 F.2d 973, 978, 154 U.S.P.Q. 92, 96 (C.C.P.A. 1967) (underline emphasis in original, italic emphasis added).

4. ***Appellants Asserted a Specific, Substantial and Credible Utility that is Sufficient to Satisfy the Utility Requirement of § 101***

The claimed subject matter is directed to polypeptides comprising the amino acid sequence of the polypeptide of SEQ ID NO:110, the amino acid sequence of the polypeptide of SEQ ID NO:110 lacking its associated signal peptide, or the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203535. Additional claimed subject matter is directed to polypeptides having at least 95% amino acid sequence identity to the amino acid sequence of the polypeptide SEQ ID NO:110, the amino acid sequence of the polypeptide of SEQ ID NO:110 lacking its associated signal peptide, or the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203535, wherein the isolated polypeptide or a fragment thereof can be used to generate an antibody which can be used to specifically detect the polypeptide of SEQ ID NO:110 in esophageal tissue samples. The polypeptide of SEQ ID NO:110 (referred to as "PRO1753 polypeptide") is encoded by the polynucleotide of SEQ ID NO:109 (also referred to as DNA68883-1691). *Specification* at ¶¶ [0135-0136]. Appellants have asserted that the claimed polypeptides are useful as diagnostic tools for cancer, particularly esophageal cancer.

In "Example 18: Tumor Versus Normal Differential Tissue Expression Distribution" Appellants disclose that the mRNA encoding PRO1753 polypeptide is more highly expressed in

esophageal tumor compared to normal esophageal tissue. *Specification* at ¶¶ [0529-0530] and accompanying tables. As explained in paragraph [0530], the differential expression of the PRO1753 mRNA was detected using the well-established technique of quantitative PCR amplification of cDNA libraries isolated from different human normal and tumor tissue samples. To ensure that equivalent amounts of nucleic acid were used in each reaction, the cDNA for β-actin was used as a control.

The specification teaches that identification of the differential expression of a PRO polypeptide-encoding mRNA in one or more tumor tissues as compared to one or more normal tissues of the same tissue type "renders the molecule useful diagnostically for the determination of the presence or absence of tumor in a subject suspected of possessing a tumor." *Specification* at ¶ [0530]. Because it is well established that changes in mRNA levels lead to changes in the level of the encoded protein, based on the teachings in the specification, one would expect the PRO1753 protein to be overexpressed in esophageal tumors. The specification states that PRO polypeptides "may also be used diagnostically for tissue typing, wherein the PRO polypeptides of the present invention may be differentially expressed in one tissue as compared to another, preferably in a diseased tissue as compared to a normal tissue of the same tissue type." *Specification* at ¶ [0336]. The specification also discloses that PRO polypeptides and polypeptides related thereto can be used to generate anti-PRO antibodies. *Id.* at ¶ [0364] and ¶ [0367]. The specification teaches that such antibodies to PRO polypeptides can be useful as diagnostic tools:

> [A]nti-PRO antibodies may be used in diagnostic assays for PRO [polypeptide], *e.g.*, detecting its expression (and in some cases, differential expression) in specific cells, tissues, or serum. Various diagnostic assay techniques known in the art may be used, such as competitive binding assays, direct or indirect sandwich assays and immunoprecipitation assays conducted in either heterogeneous or homogeneous phases. *Specification* at [0407].

Taken together, the specification clearly discloses the use of the claimed polypeptides as diagnostic tools for cancer, particularly esophageal cancer. This utility is substantial, as one of skill in the art will recognize that the diagnosis of cancer is a "real world" use; it is specific, as the diagnosis of esophageal cancer is not a utility that applies to the broad class of polypeptides;

and it is credible, as it not a utility "that could only be true if it violated a scientific principle, ...or a law of nature, or [is] wholly inconsistent with contemporary knowledge in the art." M.P.E.P. § 2107.02 III B., citing *In re Gazave*, 379 F.2d 973, 978, 154 U.S.P.Q. 92, 96 (C.C.P.A. 1967).

Because Appellants' specification contains a disclosure of utility which corresponds in scope to the claimed subject matter, the asserted utility "must be taken as sufficient to satisfy the utility requirement of § 101 for the entire claimed subject matter unless there is reason for one skilled in the art to question the objective truth of the statement of utility or its scope." *In re Langer*, 503 F.2d 1380, 1391, 183 U.S.P.Q. 288, 297 (C.C.P.A. 1974). Therefore, the burden of establishing a *prima facie* case of lack of utility rests with the PTO. *See, In re Brana*, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995) ("the PTO has the initial burden of challenging a presumptively correct assertion of utility in the disclosure").

5. ***The Examiner's Arguments***

In the first Office Action, dated February 1, 2005, the Examiner rejected the pending claims, stating "Claims 1-13 are rejected under 35 U.S.C. 101 because the claimed invention is not supported by either a specific and substantial asserted utility or a well established utility." *First Office Action* at 2. This rejection is maintained in the final Office Action. *Final Office Action* at 2.

To establish a *prima facie* showing that the claimed subject matter lacks utility, the Examiner must "provide[] evidence showing that one of ordinary skill in the art would reasonably doubt the asserted utility." *In re Brana*, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995). The Examiner has issued a first Office Action, a final Office Action, and an Advisory Action during the prosecution of the instant application. None of these papers provide any evidence that one of ordinary skill in the art would reasonably doubt the asserted utility.

The Examiner states that "it is Applicants' position that a change in gene expression establishes a significant probability that the encoded polypeptide will also be changed; that the legal standard for demonstrating utility is more likely than not, and that the standard is not absolute certainty; and, that it is more likely than not that those skilled in the art would believe

that the PRO1753 polypeptide is useful as a diagnostic tool for cancer." *Final Office Action* at 2. However, the Examiner rejects this utility. The Examiner's argument, which is repeated in whole or part throughout the final Office Action is that "[t]he examiner has cited countervailing evidence to show: firstly, that the significance or relevance of the disclosed PRO1753 mRNA expression in relation to cancer diagnosis or treatment is unknown; secondly, that protein levels [sic] are not always consistent with protein levels." *Final Office Action* at 9. While the Examiner does cite references as support for these assertions, for the reasons discussed below, the Examiner's references are lacking.

First, the Examiner cites Hu *et al.* (J. Proteome Res. (2003) 2(4):405-12) to support his statement challenging the significance and relevance of the data presented in Example 18: "Hu was cited as countervailing evidence to show that the significance of the disclosed analysis of PRO1753 mRNA expression in relation to PRO1753 polypeptide expression in cancer diagnosis or treatment is unknown." *Id.* at 7. The Examiner quotes Hu, stating "'[i]n any microarray experiment, thousands of genes may demonstrate statistically significant expression changes, but only a fraction of these may be relevant to the study' (page 405, left column, full paragraph 1). Implicit in these teachings is that there are also biologically irrelevant results, and that further research is required in order to determine which results are biologically relevant." *Final Office Action* at 8. According to the Examiner, "[o]ne skilled in the art would not know if the disclosed PRO1753 mRNA expression is significant or insignificant, relevant or irrelevant." *Id.* The Examiner also apparently relies on Wang (Trends Pharmacol. Sci., (1996) 17(8):276-9) to support his position that Appellants have not sufficiently validated the PRO1753 polypeptide. *Id.* at 9.

To support his assertion that "protein levels are not always consistent with mRNA levels and that protein levels are not predictable from the mRNA expression levels," the Examiner cites Haynes *et al.* (Electrophoresis, (1998) 19(11):1862-71) and Gygi *et al.* (Mol. and Cell. Bio., (1999). *Final Office Action* at 10-13. Finally, Hancock (J. Proteome Res., (2004) 3(4):685) is cited as being consistent with Haynes. *Id.* at 13.

The remainder of the Examiner's arguments are unsupported conclusory statements, or simply restatements of the two assertions listed above, which the Examiner repeats throughout

the Final Office Action. For example, the Examiner repeatedly makes the statement, or one nearly identical to it, that "the present specification provides no information regarding the level of expression, role or activity of the PRO1753 polypeptide in tumors." *Id.* at 3, 4, 5, 6, 7, 8, 9, 10, 14, 15, 16, 17, 18, 19 and 22. Another statement made throughout is that "[o]ne skilled in the art would not know if or how PRO1753 polypeptide expression would change in cancer." *Id.* at 3, 6, 8, 9, 11, 13, 18, and 19.

6. ***The Examiner has not established a Prima Facie case that Claims 6-8 and 11-17 lack Utility***

The above arguments do not satisfy the Examiner's burden to "provide[] evidence showing that one of ordinary skill in the art would reasonably doubt the asserted utility." *In re Brana*, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995). The Examiner has the burden of presenting "countervailing facts and reasoning sufficient to establish that a person of ordinary skill would not believe the Appellant's assertion of utility." *M.P.E.P.* at §2107.02 III.A., *citing in re Brana*, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995) ("Only after the PTO provides evidence showing that one of ordinary skill in the art would reasonably doubt the asserted utility does the burden shift to the Appellant to provide rebuttal evidence") (emphasis added). With the exception of the specific references cited by the Examiner, the Examiner's assertions are not supported by any facts, evidence, or reasoning. Instead, the Examiner simply makes unsupported assertions. As for the references cited by the Examiner, for the reasons discussed below, they do not support the Examiner's position. Therefore there is simply no evidence in the record to support the Examiner's assertion that Appellants' asserted utility is not substantial, and the invention is incomplete. Absent some substantial evidence to support his assertions, the Examiner has failed to establish a *prima facie* showing that one of skill in the art would reasonably doubt the asserted utility, and the Board should accept Appellants' disclosed utility as sufficient.

a. The data in Example 18 are sufficient to establish the asserted utility

Appellants turn first to the Examiner's arguments challenging the reliability of the data reported in Example 18. The Examiner argues that Example 18 and the first declaration of Mr. Grimaldi are insufficient to overcome the utility rejection. In particular, the Examiner cites Hu *et al.* as countervailing evidence to show that the significance of the disclosed analysis of PRO1753 mRNA expression in relation to PRO1753 polypeptide expression in cancer diagnosis or treatment is unknown. *Final Office Action* at 7-8. According to the Examiner, "[o]ne skilled in the art would not know if the disclosed PRO1753 mRNA expression is significant or insignificant, relevant or irrelevant." *Id.*

As an initial matter, Appellants note that, other than the Examiner's assertions based on Hu *et al.*, the remainder of the objections to the data reported in Example 18 are not supported by any evidence or reasoning as to why the data are insufficient, and therefore they cannot establish a *prima facie* case. *See In re Brana*, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995) ("Only after the PTO provides evidence showing that one of ordinary skill in the art would reasonably doubt the asserted utility does the burden shift to the Appellant to provide rebuttal evidence.") (emphasis added). Despite this deficiency, Appellants address the Examiner's objections below.

The gene expression data in Example 18 of the specification show that the mRNA associated with protein PRO1753 was more highly expressed in esophageal tumor compared to normal esophageal tissue. *See Specification* at ¶ [0530] and accompanying tables. Gene expression was analyzed using standard quantitative PCR amplification reactions of cDNA libraries isolated from different human tumor and normal human tissue samples. *Id.* It is well known in the art that the number of copies of a particular cDNA in the cDNA library is determined by the number of copies of the corresponding mRNA in the sample. Therefore, the cDNA libraries can be used to determine the level of expression of the corresponding mRNA in the tissue.

Appellants have asserted that identification of the differential expression of the PRO1753 polypeptide-encoding gene in tumor tissue compared to the corresponding normal tissue renders antibodies against the polypeptide useful as a diagnostic tool for the determination of the

presence or absence of tumor. *Id.* In support of this asserted utility, Appellants submitted on April 29, 2005 as Exhibit 1 to their Amendment and Response to Office Action a first Declaration of J. Christopher Grimaldi, an expert in the field of cancer biology. This declaration explains the importance of the data in Example 18, and how differential gene and protein expression studies are used to differentiate between normal and tumor tissue. *See First Grimaldi Declaration.*

In paragraphs 6 and 7, Mr. Grimaldi explains that the semi-quantitative analysis employed to generate the data of Example 18 is sufficient to determine if a gene is over- or under-expressed in tumor cells compared to corresponding normal tissue. He states that any visually detectable difference seen between two samples is indicative of at least a two-fold difference in cDNA between the tumor tissue and the counterpart normal tissue. He also states that the results of the gene expression studies indicate that the genes of interest "can be used to differentiate tumor from normal." He explains that, contrary to the PTO's assertions, "[t]he precise levels of gene expression are irrelevant; what matters is that there is a relative difference in expression between normal tissue and tumor tissue." *First Grimaldi Declaration* at ¶ 7.

This declaration makes clear that since it is the relative level of expression between normal tissue and suspected cancerous tissue that is important, how high the level of expression in normal tissue is, is irrelevant. As to the Examiner's questions about the reliability of the results, Appellants employed standard techniques which are well-known and accepted by those of skill in the art. Mr. Grimaldi states that if a difference is detected using these techniques, "this indicates that the gene and its corresponding polypeptide and antibodies against the polypeptide are useful for diagnostic purposes..." *Id.* The Grimaldi Declaration also states that the samples are pooled samples of normal and tumor tissue, and therefore are more reliable than individual samples. *Id.* at ¶ 5. Appellants are not required to prove utility to a statistical certainty, only that it is more likely than not true. *See Nelson v. Bowler*, 626 F.2d 853, 856-57, 206 U.S.P.Q. 881, 883-84 (C.C.P.A. 1980) (reversing the Board and rejecting an argument that evidence of utility was insufficient because it was not statistically significant). Therefore, whether the results are statistically certain or not is irrelevant to establishing the asserted utility. Thus, it is the

uncontested opinion of an expert in the field that the results are reliable enough to indicate that the claimed antibodies are useful as diagnostic tools.

Hu is cited by the Examiner as countervailing evidence to show that the significance of the disclosed analysis of PRO1753 mRNA expression in relation to PRO1753 polypeptide expression in cancer diagnosis or treatment is unknown. *Final Office Action* at 7-8. The Examiner quotes Hu, stating "'[i]n any microarray experiment, thousands of genes may demonstrate statistically significant expression changes, but only a fraction of these may be relevant to the study' (page 405, left column, full paragraph 1). Implicit in these teachings is that there are also biologically irrelevant results, and that further research is required in order to determine which results are biologically relevant." *Id.* at 8. According to the Examiner, "[o]ne skilled in the art would not know if the disclosed PRO1753 mRNA expression is significant or insignificant, relevant or irrelevant." *Id.* at 8.

In Hu, the researchers used an automated literature-mining tool to summarize and estimate the relative strengths of all human gene-disease relationships published on Medline. They then generated a microarray expression dataset comparing breast cancer and normal breast tissue. Using their data-mining tool, they looked for a correlation between the strength of the literature association between the gene and breast cancer, and the magnitude of the difference in expression level. They report that for genes displaying a 5-fold change or less in tumors compared to normal, there was no evidence of a correlation between altered gene expression and a *known* role in the disease. *See Hu* at 411. However, among genes with a 10-fold or more change in expression level, there was a strong correlation between expression level and a *published* role in the disease. *Id.* at 412. Importantly, Hu reports that the observed correlation was only found among estrogen receptor-positive tumors, not ER-negative tumors. *Id.*

The general findings of Hu are not surprising – one would expect that genes with the greatest change in expression in a disease would be the first targets of research, and therefore have the strongest known relationship to the disease as measured by the number of publications reporting a connection with the disease. The correlation reported in Hu only indicates that the greater the change in expression level, the more likely it is that there is a *published* or *known* role for the gene in the disease, as found by their automated literature-mining software. Thus, Hu's

results merely reflect a bias in the literature toward studying the most prominent targets, and reflect nothing regarding the ability of a gene that is 2-fold or more differentially expressed in tumors to serve as a disease marker.

Hu acknowledges the shortcomings of this method in explaining the disparity in Hu's findings for ER-negative versus ER-positive tumors: Hu attributes the "bias in the literature" toward the more prevalent ER-positive tumors as the explanation for the lack of any correlation between number of publications and gene expression levels in less-prevalent (and, therefore, less studied) ER-negative tumors. *Id.* Because of this intrinsic bias, Hu's methodology is unlikely to ever note a correlation of a disease with less differentially-expressed genes and their corresponding proteins, regardless of whether or not an actual relationship between the disease and less differentially-expressed genes exists. Accordingly, Hu's methodology yields results that provide little or no information regarding biological significance of genes with less than 5-fold expression change in disease. Nowhere in Hu does it say that a lack of correlation in their study means that genes with a less than five-fold change in level of expression in cancer cannot serve as a molecular marker of cancer.

Appellants submit that a lack of known role for PRO1753 in cancer does not prevent it, or antibodies which bind to it, from being useful as diagnostic tools for cancer. There is a difference between use of a gene for distinguishing between tumor and normal tissue on the one hand, and establishing a role for the gene in cancer on the other. Genes with lower levels of change in expression may or may not be the most important genes in causing the disease, but the genes can still show a consistent and measurable change in expression. While such genes may or may not be good targets for further research, the encoded polypeptides and antibodies which bind to them can nonetheless be used as diagnostic tools. Thus, Hu does not refute the Appellants' assertion that the claimed polypeptides can be used as a cancer diagnostic tools because the polypeptide is differentially expressed in certain tumors.

Contrary to the Examiner's assertion that one must know what role a gene or polypeptide plays in cancer for it have utility, the PTO's own written policies recognize that the utility of a nucleic acid does not depend on the function of the encoded gene product. The Utility Examination Guidelines published on January 5, 2001 state: "In addition, the utility of a claimed

DNA does not necessarily depend on the function of the encoded gene product. A claimed DNA may have a specific and substantial utility because, e.g. it hybridizes near a disease-associated gene or it has a gene regulating activity." (Federal Register, Volume 66, page 1095, Comment 14). Similarly, here the disclosed nucleic acids, as well as the encoded polypeptides and related antibodies, are useful for determining whether an individual has cancer regardless of whether or not they are the cause of the cancer.

The position of the Examiner requiring a known role for PRO1753 in cancer for utility is also inconsistent with the analogous standard for therapeutic utility of a compound where "the mere identification of a pharmacological activity of a compound that is relevant to an asserted pharmacological use provides an 'immediate benefit to the public' and thus satisfies the utility requirement." *M.P.E.P.* §2701.01 (emphasis original). Here, the mere identification of altered expression in tumors is relevant to diagnosis of tumors, and, therefore, provides an immediate benefit to the public.

In addition, the Grimaldi Declaration provides further facts relating to Example 18 by stating that "DNA libraries used in the gene expression studies were made from pooled samples of normal and of tumor tissues." *First Grimaldi Declaration* at ¶ 5. Mr. Grimaldi then explains the significance of this fact by stating that "[d]ata from pooled samples is more likely to be accurate than data obtained from a sample from a single individual." *Id.* Mr. Grimaldi further states that differential expression observed from such pooled samples indicates that the differentially expressed gene is "useful for diagnostic purposes." *Id.* at ¶ 7.

The Examiner bases his dismissal of the Grimaldi Declaration on the Hu reference, stating "'[i]n any microarray experiment, thousands of genes may demonstrate statistically significant expression changes, but only a fraction of these may be relevant to the study' (page 405, left column, full paragraph 1). Implicit in these teachings is that there are also biologically irrelevant results, and that further research is required in order to determine which results are biologically relevant." *Final Office Action* at 8. According to the Examiner, "[o]ne skilled in the art would not know if the disclosed PRO1753 mRNA expression is significant or insignificant, relevant or irrelevant." *Id.* at 8.

While Hu does discuss "biologically relevant results," Hu does not discuss or mention "diagnostic relevance" or identify a threshold differential expression level needed to establish the relevance of a diagnostic marker. This is because Hu merely attempted to establish a relationship between differential expression levels and a published or known role for a gene in disease, as found by automated literature-mining software. Hu did not examine or discuss the use of differentially expressed genes as biological markers or diagnostic tools.

Moreover, in dismissing the first Grimaldi Declaration, the Examiner fails to give weight to the fact that Example 18 discloses differential expression data from pooled samples, and that "[d]ata from pooled samples is more likely to be accurate than data obtained from a sample from a single individual." *First Grimaldi Declaration* at ¶ 5. The Examiner provides no basis for dismissing these statements by Grimaldi. Hu cannot support the Examiner's dismissal of the Grimaldi Declaration because Hu is silent regarding the use of differentially expressed genes as diagnostic tools in general, and the reliability of pooled samples in particular. Hu says nothing about whether or not differential expression in pooled samples is susceptible to disease-independent differences between samples. Hu provides no reason to expect that differential expression in pooled samples is attributable to disease-independent differences between samples. Thus, Hu does not provide a basis for dismissing the first Grimaldi Declaration. The Examiner's unsupported statement that "[o]ne skilled in the art would not know if the disclosed PRO1753 mRNA expression is significant or insignificant, relevant or irrelevant" is addressed by Grimaldi's statements that the pooled samples used are reliable, and that the reported differential expression is sufficient to differentiate tumor tissue from normal tissue. *First Grimaldi Declaration* at ¶¶ 5, 7.

The data in Example 18 and the first Grimaldi Declaration are therefore sufficient to establish the asserted utility, and the Examiner has not rebutted the presumption of utility that the Appellants' application is afforded. Mr. Grimaldi is an expert in the field who conducted or supervised the experiments at issue. His declaration is based on personal knowledge of the relevant facts at issue. Appellants' have reminded the Examiner that "Office personnel must accept an opinion from a qualified expert that is based upon relevant facts whose accuracy is not being questioned." *M.P.E.P.* § 2107 (emphasis added). In addition, declarations relating to

issues of fact should not be summarily dismissed as "opinions" without an adequate explanation of how the declaration fails to rebut the Examiner's position. *See in re Alton* 76 F.3d 1168 (Fed. Cir. 1996).

The Examiner has offered no reason or evidence to reject either the underlying data or the conclusions of the Grimaldi Declaration, other than the cited Hu reference which is not contrary to Appellants' assertion. It is important to note that nowhere in Hu does it say that a lack of correlation in their study means that genes with a less than five-fold change in level of expression in cancer cannot serve as a molecular marker of cancer. Likewise, Hu is silent regarding the reliability of pooled samples. As Hu is the only evidence cited by the Examiner to support his assertion that the data in Example 18 are unreliable and insufficient to support Appellants' asserted utility, the Examiner is left with no basis to question the conclusion of the first Grimaldi Declaration or Appellants' assertions. Therefore, the Examiner must accept Mr. Grimaldi's opinion with regard to his statement that the genes of interest "can be used to differentiate tumor from normal."

In addition to Hu *et al.*, the Examiner also relies on Wang to support his argument that Appellants have not sufficiently validated the role of PRO1753 gene in disease processes. *First Office Action* at 5; *Final Office Action* at 8-9. In Wang, the authors outline a strategy for using mRNA differential display for the discovery of novel pharmacological targets. They state that the use of mRNA differential display for the isolation of novel genes associated with disease processes will no doubt facilitate discovery of novel pharmaceutical targets. *See Wang* at 279. However, they state that it is the first of many steps in the process, and that characterization of the functions of the gene as well as validation of the importance of the gene in disease processes is required. *Id.*

Nothing in Wang is contrary to Appellants' assertion that differentially expressed genes and polypeptides can be used as molecular markers of cancer. Wang speaks only of the additional steps needed to develop a novel pharmacological target. Contrary to the Examiner's assertions, nowhere does Wang teach that differentially expressed genes and polypeptides cannot be used as molecular markers for cancer. Thus, Wang does not refute the Appellants' assertion

that the PRO1753 gene, polypeptide, and antibodies can be used as a cancer diagnostic tools because they are differentially expressed in esophageal tumors.

In the Final Office Action, the Examiner responds to Appellants' arguments by stating that "[t]he examiner can accept, for arguments sake, that gene expression analysis, tools or techniques have the potential for the discovery of new diagnostic or therapeutic targets. However, in the present case no information is provided in the differential analysis of PRO1753 mRNA expression regarding the level of expression, activity, or role in cancer of the PRO1753 polypeptide." *Final Office Action* at 9. According to the Examiner, further characterization of PRO1753 polypeptide is part of the act of invention and until it has been undertaken, Appellants' invention is incomplete. The Examiner disagrees with Appellants' assertion that they have completed the equivalent of steps 1-9 of the ten steps outlined by Wang in Figure 1 because "the present specification does not provide any information regarding the target validation of the PRO1753 polypeptide." *Id.* Thus, the Examiner relies on Wang for the requirements that: (1) the specification provide information regarding expression, activity, or role in cancer of the PRO1753 polypeptide; and (2) the PRO1753 polypeptide must also have undergone validation for the importance of the gene in disease processes.

Appellants submit that the Examiner's response fails to address Appellants' assertion that Wang does not offer any support for the Examiner's position since the reference does not address the use of a differentially expressed gene or polypeptide as a diagnostic tool.

Nothing in Wang supports a requirement for information regarding the level of expression, activity, or role in cancer for a polypeptide before it is useful as a diagnostic tool since Wang does not address the development of novel diagnostic markers for cancer. In addition, as is explained in detail below, the understanding in the art is that there is a reasonable correlation between changes in gene expression and changes in the expression of the corresponding protein. Thus, contrary to the Examiner's repeated assertions, Appellants have provided information regarding the expression of the PRO1753 polypeptide, as one of skill in the art would know that PRO1753 polypeptide expression is expected to be increased in esophageal tumor compared to normal esophageal tissue.

Furthermore, the Examiner's requirement that the PRO1753 polypeptide must also have undergone validation for the importance of the gene in disease processes is irrelevant to Appellants' asserted utility of the claimed polypeptide in the diagnosis of cancer. As discussed above in relation to the Hu *et al.* reference, it is not necessary to determine the importance of a gene or the encoded polypeptide in a disease process in order to establish utility for the polypeptide as a diagnostic tool for that disease. Therefore, the Examiner's requirement of validation in order to establish utility is not germane to Appellants' asserted utility.

Because Wang does not discuss the requirements necessary for developing a novel molecular marker of cancer for use as a diagnostic tool, the Examiner cannot rely on Wang to reject Appellants' asserted utility. Wang is not relevant, and nothing in Wang would lead one skilled in the art to question Appellants' asserted utility.

In conclusion, Appellants submit that the evidence reported in Example 18, which are supported by the first Grimaldi Declaration, establish that there is at least a two-fold difference in PRO1753 mRNA between esophageal tumor and normal esophageal tissue, and that the data in Example 18 are significant and reliable enough to use PRO1753 to distinguish esophageal tumor tissue from its normal tissue counterpart. The Examiner has not offered any significant arguments or evidence to the contrary, and therefore has not established a *prima facie* case that one of skill in the art would reasonably doubt the asserted utility of the PRO1753 mRNA, or, as explained below, the PRO1753 polypeptide.

b. The references cited by the Examiner do not refute Appellants' assertion that a change in mRNA levels leads to a corresponding change in the level of the encoded protein

Appellants turn next to Examiner's second argument that "protein levels are not always consistent with mRNA levels" and therefore "the skilled artisan would not know if or how PRO1753 polypeptide expression would change in cancer." *Final Office Action* at 3. The Examiner cites Haynes and Gygi as providing evidence that protein expression levels are not predictable from the mRNA expression levels. *Id.* at 10-13. The Examiner also argues that Hancock is consistent with Haynes. *Id.* at 13. For the reasons discussed below, none of these

references are contrary to Appellants' assertion that generally speaking, changes in mRNA levels lead to corresponding changes in the level of polypeptide.

Haynes studied whether there is a correlation between the level of mRNA expression and the level of protein expression for 80 selected genes from yeast. The genes were selected because they constituted a relatively homogeneous group with respect to predicted half-life and expression level of the protein products. *See Haynes* at 1863. Haynes did not examine whether a **change** in transcript level for a particular gene led to a **change** in the level of expression of the corresponding protein. Instead, Haynes determined whether the steady-state transcript level correlated with the steady-state level of the corresponding protein based on an analysis comparing levels of mRNA and protein across 80 different genes. The Examiner also relies on Gygi *et al.*, a study on which the Haynes reference is based. Like Haynes, the Gygi reference looked at static levels of mRNA across different genes, not changes in mRNA levels for a single gene.

According to the Examiner, Haynes and Gygi, considered as a whole, teach "that protein levels cannot be accurately predicted from the level of the corresponding mRNA transcript and that there are numerous levels of control of protein synthesis, degradation, processing and modification, which are only apparent by direct protein analysis." *Id.* at 12-13. Based on these teachings, the Examiner concludes that "the skilled artisan would not know if a change in mRNA level is associated with a corresponding change in protein levels and that the skilled artisan would have a legitimate basis to doubt the utility of the PRO1753 polypeptide because the skilled artisan would not know if or how PRO1753 polypeptide expression would change in cancer." *Id.* at 13.

Appellants emphasize that neither Haynes nor Gygi looked at whether a change in transcript level for a particular gene led to a change in the level of expression of the corresponding protein – neither reference examined differential expression of mRNA and proteins. Instead, these references considered whether the steady-state transcript level correlated with the steady-state level of the corresponding protein by comparing mRNA and protein levels across many different genes. These references disclose that similar mRNA levels for different genes did not universally result in equivalent protein levels for the different gene products, and

that similar protein levels for different gene products did not universally result from equivalent mRNA levels for the different genes. These results are expected, since there are many factors that determine translation efficiency for a given transcript, or the half-life of the encoded protein. Not surprisingly, based on these results, Haynes and Gygi concluded that protein levels cannot always be accurately predicted from the level of the corresponding mRNA transcript when looking at the steady-state level of transcripts across different genes.

Importantly, Haynes and Gygi did not say that for a single gene, changes in the level of mRNA transcript are not positively correlated with changes in the level of protein expression. Appellants have asserted that increasing or decreasing the level of mRNA for the same gene leads to a increase or decrease for the encoded protein – that is to say that differential mRNA expression leads to differential protein expression. Haynes and Gygi did not study this issue and say absolutely nothing about it. One cannot look at the level of mRNA across several different genes to investigate whether a change in the level of mRNA for a particular gene leads to a change in the level of protein for that gene.

Haynes and Gygi have nothing to do with changes in mRNA levels resulting in changes in protein levels because they only examined static levels of mRNA and protein compared across different genes. Thus, when these references assert that protein levels cannot be accurately predicted from mRNA levels, they are referring to the finding that similar steady-state levels of mRNA for different genes can lead to the different levels of protein for the corresponding gene products, and that the similar steady-state level of protein for different genes can result from different levels of the corresponding mRNAs. This is completely unrelated to the assertion that a change in mRNA levels leads to a corresponding change in the level of the encoded protein.

Neither Haynes nor Gygi provide any insight whatsoever into the affects on protein level caused by a change in the encoding mRNA level. The Examiner has not provided any evidence or reasons which support extending the findings of Haynes and Gygi regarding steady-state levels of mRNA for different genes to reject Appellants' assertions regarding changes in mRNA levels for the same gene. While Applicants acknowledge that protein levels are regulated at numerous levels, including post-translational mechanisms, Haynes and Gygi only provide insight on how these mechanisms lead to differences in protein levels across different genes. They do not

address the Appellants' assertion that generally, changes in mRNA level for a particular gene lead to changes in the level of the encoded protein. Since this issue is not addressed in Haynes and Gygi, these references offer no support for the Examiner's rejection of Appellants' asserted utility.

The final reference cited by the Examiner is a single page editorial by Hancock, in which the author states that "markers generated by proteomics are not always consistent with markers that are generated from expression profiling." *Hancock* at 685. This single sentence expressing the unsupported opinion of one scientist is relied on by the Examiner to support the assertion that "transcript levels are not always correlated with protein levels." *First Office Action* at 4.

Read in context, it is not clear that Hancock's statement supports the Examiner's assertion. The point of Hancock's editorial is that proteomics is a developing and untested technology. After the statement quoted above, the author continues: "This Editor believes that proteomics is at too early a stage for this new technology to have generated a quality list of [bio]markers." *Hancock* at 685. Thus, it appears that rather than suggesting that mRNA levels are not always correlated with protein levels, Hancock is instead arguing that proteomics has not developed sufficiently to be a reliable method of generating biomarkers. In addition, as with Haynes and Gygi, Hancock says nothing about Appellants' assertion that increasing or decreasing the level of mRNA for a gene leads to an increase or decrease in the corresponding protein.

In response to Appellants' arguments, the Examiner states that Hancock is consistent with Haynes. *Final Office Action* at 13. Specifically, the Examiner asserts that Hancock supports his assertion that the analysis of protein products is essential because protein expression levels are not predictable from the mRNA expression levels. *Id.* The Examiner asserts that one skilled in the art would be required to do further research in order to determine whether or not the PRO1753 polypeptide levels changed significantly in the tumor samples. *Id.*

Appellants submit that this response does not address Appellants' criticism of Hancock. As discussed above, Haynes does not support the Examiner's position as it does not address Appellants' assertion that changes in mRNA levels lead to changes in the level of the corresponding protein. Thus, even if Hancock is consistent with Haynes, it does not provide any support for the Examiner's rejection of Applicants' asserted utility.

c. ***Conclusion – Examiner has failed to establish a prima facie case that one of skill in the art would doubt Appellants' asserted utility***

The Examiner has questioned the sufficiency, reliability and significance of the data reported in Example 18 as well as the supporting first Grimaldi Declaration, relying on Hu *et al.* for support. The Examiner also apparently continues to rely on Wang to support his position that Appellants have not sufficiently validated the PRO1753 polypeptide. To support his assertion that polypeptide levels cannot be accurately predicted from mRNA levels, the Examiner relies on Haynes *et al.*, Gygi *et al.*, and Hancock. Appellants submit that these arguments and supporting references fail to establish a prima facie case that one skilled in the art would reasonably doubt Appellants' asserted utility.

First, Appellants have shown that the data in Example 18 are sufficiently significant and reliable to establish that PRO1753 is useful as a cancer diagnostic tool. This assertion is supported by the first Grimaldi Declaration. The Examiner has not provided any substantial reason or evidence for one of skill in the art to doubt the reliability or usefulness of Example 18, or the facts and conclusions in the first Grimaldi Declaration. In particular, Appellants maintain Hu *et al.* does not contradict Appellants' position because Hu *et al.* focuses on the role of polypeptides in cancer, not whether differentially expressed mRNAs can be used as diagnostic markers. In addition, Hu does not address the reliability of pooled samples, and therefore cannot be relied on to counter the expert opinions expressed in the first Grimaldi Declaration that the differential expression of PRO1753 makes it useful as a diagnostic tool. Similarly, Appellants maintain that Wang relates to the development and validation of novel pharmacological targets, not diagnostic tools. Therefore, like Hu, Wang is simply not relevant, and does not support the rejection of Appellants' asserted utility.

The final references relied on by the Examiner are Haynes *et al.*, Gygi *et al.*, and Hancock. Appellants have shown that the Haynes and Gygi references are simply not relevant to the issue of whether a change in mRNA levels leads to a corresponding change in the level of the encoded protein since the references examined steady-state levels of mRNA and protein across different genes. Furthermore, rather than suggesting that mRNA levels are not always correlated

with protein levels, Hancock argues that proteomics has not developed sufficiently to be a reliable method of generating biomarkers. To the extent that it is consistent with Haynes, Hancock is also irrelevant.

Taken together, the Examiner's arguments are not sufficient to satisfy the Examiner's burden to "provide[] evidence showing that one of ordinary skill in the art would reasonably doubt the asserted utility." *In re Brana*, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995). The Examiner's arguments are largely conclusory statements which are not supported by any substantial evidence or reasoning which explains why one of ordinary skill in the art would reasonably doubt the asserted utility. Therefore, the Board should accept the Appellants' disclosure of utility. *See Ex parte Rubin*, 5 U.S.P.Q. 2d 1461 (Bd. Pat. App. & Interf. 1987) ("There is no factual support in this record for the examiner's questioning of the denaturation test reported in the specification. ... No reason to doubt 'the objective truth' of the asserted utility having been advanced by the examiner, we accept appellant's disclosure of utility corresponding in scope to the claimed subject matter.").

7. ***Appellants have provided Sufficient Rebuttal Evidence of Utility***

"Only after the PTO provides evidence showing that one of ordinary skill in the art would reasonably doubt the asserted utility does the burden shift to the applicant to provide rebuttal evidence." I*n re Brana*, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995). The rebuttal evidence must be sufficient such that when it is considered as a whole, it is more likely than not that the asserted utility is true. *See In re Oetiker*, 977 F.2d 1443, 1445, 24 U.S.P.Q.2d 1443, 1444 (Fed. Cir. 1992) (stating that the evidentiary standard to be used throughout *ex parte* examination in setting forth a rejection is a preponderance of the evidence, or "more likely than not" standard). The M.P.E.P. summarizes the standard of proof required:

> [T]he applicant does not have to provide evidence sufficient to establish that an asserted utility is true "beyond a reasonable doubt." Nor must the applicant provide evidence such that it establishes an asserted utility as a matter of statistical certainty. Instead, evidence will be sufficient if, considered as a whole, it leads a person of ordinary skill in the art to conclude that the asserted utility is more likely than not true. *M.P.E.P.* § 2107.02, part VII (emphasis in original, citations omitted).

Appellants remind the Board that the Federal Circuit has stated that the standard for satisfying the utility requirement is a low one: "The threshold of utility is not high: An invention is 'useful' under section 101 if it is capable of providing some identifiable benefit." *Juicy Whip, Inc. v. Orange Bang, Inc.*, 185 F.3d 1364, 1366, 51 U.S.P.Q. 2d 1700 (Fed. Cir. 1999).

Even if the Examiner has satisfied his burden of presenting a *prima facie* case of lack of utility, Appellants have supplied more than enough rebuttal evidence, such that when considered as a whole, one of skill in the art would conclude that the asserted utility is more likely than not true. As discussed in detail below, Appellants have provided sufficient evidence that the gene encoding the PRO1753 polypeptide is differentially expressed in esophageal tumors and can therefore be used as a diagnostic tool. In addition, Appellants have shown that it is well established in the art that there is a reasonable correlation between changes in mRNA level and changes in the corresponding protein level such that one of skill in the art would believe that the PRO1753 polypeptide is also differentially expressed in esophageal cancers. Therefore, considering the evidence as a whole, one of skill in the art would believe that it is more likely than not that the claimed polypeptides are useful as diagnostic tools for cancer, particularly esophageal cancer.

a. Appellants have established that the gene encoding the PRO1753 polypeptide is differentially expressed in certain cancers

As discussed above, the Examiner has not provided any relevant evidence or reasoning to challenge the reliability and significance of the data in Example 18 which reports that the mRNA for PRO1753 is more highly expressed in esophageal tumor tissue compared to normal esophageal tissue. In contrast to this complete lack of relevant evidence on the part of the Examiner, Appellants have submitted the first Grimaldi Declaration. That declaration establishes that it is the opinion of an expert in the field who has personal knowledge of the facts surrounding Example 18 that there is at least a two-fold difference in mRNA for PRO1753 between the tumor tissue and the counterpart normal tissue, and that the PRO1753 genes, polypeptides and antibodies are useful for differentiating tumor tissue from normal tissue. As discussed above, Hu *et al.* and Wang *et al.* are not relevant to determining whether differentially

expressed genes can be used as diagnostic tools for cancer. Therefore, the Examiner has not provided any relevant evidence or reasoning to challenge the facts and conclusions of the first Grimaldi Declaration in support of Example 18.

Given the disclosure of Example 18 and the supporting first Grimaldi Declaration on the one hand, and the complete lack of any relevant evidence on the other, it is clear that considering the evidence as a whole, one of skill in the art would conclude that it is more likely than not that the PRO1753 gene is differentially expressed in esophageal tumor tissue compared to its normal tissue counterpart such that it is useful as a diagnostic tool to distinguish tumor tissue from normal tissue.

As Appellants explain below, it is more likely than not that the PRO1753 polypeptide is also differentially expressed in esophageal tumors, and can therefore be used to distinguish tumor tissue from normal tissue.

b. Appellants have established that generally there is a correlation between changes in mRNA expression levels and changes in the expression level of the encoded protein

Appellants next turn to the second portion of their argument in support of their asserted utility – that it is well-established in the art that in most cases a change in the level of mRNA for a particular protein leads to a corresponding change in the level of the encoded protein. Given Appellants' evidence of increased expression of the mRNA for the PRO1753 polypeptide in esophageal tumor compared to normal esophageal tissue, it is more likely than not that the PRO1753 polypeptide is likewise differentially expressed, and therefore the claimed polypeptides are useful as diagnostic tools, particularly for esophageal tumors.

In support of the assertion that changes in mRNA are positively correlated to changes in protein levels, Appellants submitted a second Declaration by J. Christopher Grimaldi, an expert in the field of cancer biology (originally submitted on April 29, 2005 as Exhibit 5 with the Appellants' Amendment and Response to Office Action). As stated in paragraph 5 of the declaration, "Those who work in this field are well aware that in the vast majority of cases, when a gene is over-expressed...the gene product or polypeptide will also be over-expressed.... This

same principal applies to gene under-expression." *Second Grimaldi Declaration* at ¶ 5. Further, "increased mRNA expression is expected to result in increased polypeptide expression, and the detection of decreased mRNA expression is expected to result in decreased polypeptide expression." *Id.*

Appellants also submitted the declaration of Paul Polakis, Ph.D. an expert in the field of cancer biology (attached as Exhibit 6 to Appellants' Amendment and Response to Office Action submitted on April 29, 2005). As stated in paragraph 6 of his declaration:

> Based on my own experience accumulated in more than 20 years of research, including the data discussed in paragraphs 4 and 5 above [showing a positive correlation between mRNA levels and encoded protein levels in the vast majority of cases studied in relation to the present invention] and my knowledge of the relevant scientific literature, it is my considered scientific opinion that for human genes, an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell. In fact, it remains a central dogma in molecular biology that increased mRNA levels are predictive of corresponding increased levels of the encoded protein. *Polakis Declaration* at ¶ 6 (emphasis added).

Dr. Polakis acknowledges that there are published cases where such a correlation does not exist, but states that it is his opinion, based on over 20 years of scientific research, that "such reports are exceptions to the commonly understood general rule that increased mRNA levels are predictive of corresponding increased levels of the encoded protein." *Polakis Declaration* at ¶ 6.

The statements of Grimaldi and Polakis are supported by the teachings in Molecular Biology of the Cell, a leading textbook in the field (Bruce Alberts, *et al.*, Molecular Biology of the Cell (3rd ed. 1994) (submitted on April 29, 2005 with Appellants' Amendment and Response to Office Action as Exhibit 7, hereinafter "Cell 3rd") and (4th ed. 2002) (submitted on April 29, 2005 with Appellants' Amendment and Response to Office Action as Exhibit 8, hereinafter "Cell 4th")). Figure 9-2 of Cell 3rd shows the steps at which eukaryotic gene expression can be controlled. The first step depicted is transcriptional control. Cell 3rd provides that "[f]or most genes transcriptional controls are paramount. This makes sense because, of all the possible control points illustrated in Figure 9-2, only transcriptional control ensures that no superfluous intermediates are synthesized." *Cell 3rd* at 403 (emphasis added). In addition, the text states that

"Although controls on the initiation of gene transcription are the predominant form of regulation for most genes, other controls can act later in the pathway from RNA to protein to modulate the amount of gene product that is made." *Cell 3rd* at 453 (emphasis added). Thus, as established in Cell 3rd, the predominant mechanism for regulating the amount of protein produced is by regulating transcription.

In Cell 4th, Figure 6-3 on page 302 illustrates the basic principle that there is a correlation between increased gene expression and increased protein expression. The accompanying text states that "a cell can change (or regulate) the expression of each of its genes according to the needs of the moment – *most obviously by controlling the production of its mRNA*." *Cell 4th* at 302 (emphasis added). Similarly, Figure 6-90 on page 364 of Cell 4th illustrates the path from gene to protein. The accompanying text states that while potentially each step can be regulated by the cell, "the initiation of transcription is the most common point for a cell to regulate the expression of each of its genes." *Cell 4th* at 364 (emphasis added). This point is repeated on page 379, where the authors state that of all the possible points for regulating protein expression, "[f]or most genes transcriptional controls are paramount." *Cell 4th* at 379 (emphasis added).

Further support for Appellants' position can be found in the textbook, Genes VI, (Benjamin Lewin, Genes VI (1997)) (submitted on April 29, 2005 with Appellants' Amendment and Response to Office Action as Exhibit 9) which states "having acknowledged that control of gene expression can occur at multiple stages, and that production of RNA cannot inevitably be equated with production of protein, it is clear that the overwhelming majority of regulatory events occur at the initiation of transcription." *Genes VI* at 847-848 (emphasis added).

Additional support is also found in Zhigang *et al.*, World Journal of Surgical Oncology 2:13, 2004 (submitted on April 29, 2005 with Appellants' Amendment and Response to Office Action as Exhibit 10). Zhigang studied the expression of prostate stem cell antigen (PSCA) protein and mRNA to validate it as a potential molecular target for diagnosis and treatment of human prostate cancer. The data showed "a high degree of correlation between PSCA protein and mRNA expression" *Zhigang* at 4. Of the samples tested, 81 out of 87 showed a high degree of correlation between mRNA expression and protein expression. The authors conclude that "it is demonstrated that PSCA protein and mRNA overexpressed in human prostate cancer, and that

the increased protein level of PSCA was resulted from the upregulated transcription of its mRNA." *Id.* at 6. Even though the correlation between mRNA expression and protein expression occurred in 93% of the samples tested, not 100%, the authors state that "PSCA may be a promising molecular marker for the clinical prognosis of human Pca and a valuable target for diagnosis and therapy of this tumor." *Id.* at 7.

Further, Meric *et al.*, Molecular Cancer Therapeutics, vol. 1, 971-979 (2002), (submitted on April 29, 2005 with Appellants' Amendment and Response to Office Action as Exhibit 11), states the following:

> The **fundamental principle** of molecular therapeutics in cancer is to exploit the differences in gene expression between cancer cells and normal cells...[M]ost efforts have concentrated on identifying differences in gene expression at the level of mRNA, which can be attributable to either DNA amplification or to differences in transcription. *Meric et al.* at 971 (emphasis added).

Exploiting differences in gene expression between cancer cells and normal cells would not be a "fundamental principle" of molecular cancer therapeutics if there were no significant correlation between gene expression and protein levels. Stated another way, changes in mRNA without corresponding changes in protein levels would have little or no effect on cellular biology, and those of skill in the art would have no reason to examine the differences in gene expression at the mRNA level without such a correlation. However, as one of skill in the art recognizes, there is a strong correlation between changes in mRNA and changes in protein level. It is because of this strong correlation that it remains a "fundamental principle" of molecular therapeutics in cancer to look at changes in mRNA level.

Together, the declarations of Grimaldi and Polakis, the accompanying references, and the excerpts and references discussed above all establish that the accepted understanding in the art is that there is a reasonable correlation between changes in gene expression and changes in the level of the encoded protein. In contrast to this substantial amount of evidence supporting Appellants' position, the Examiner has cited two references, Haynes *et al.* and Gygi *et al.* which examined the same data, as well as an editorial by Hancock. However, as discussed above, Haynes *et al.*, Gygi *et al.* and Hancock are not relevant to the issue of whether a change in mRNA levels leads to a change in the level of the corresponding protein since both Haynes and Gygi looked only at

steady-state mRNA and protein levels across different genes. It is clear that when considered as a whole, the preponderance of the evidence clearly weighs in favor of Appellants.

Appellants have presented sufficient evidence to establish that the mRNA for PRO1753 is differentially expressed in esophageal tumor tissue compared to its normal tissue counterpart, and that it is more likely than not that this leads to differential expression of the PRO1753 polypeptide. This makes the claimed polypeptides useful for diagnosing cancer, particularly esophageal cancer. Given the overwhelming amount of evidence in support of Appellants' position, and the absence of any evidence in support of the Examiner's position, when considered as a whole the evidence leads a person of ordinary skill in the art to conclude that the asserted utility is more likely than not true.

c. The asserted utility is specific

Finally, Appellants address the PTO's assertion that the asserted utilities are not specific to the claimed polypeptides.

Specific utility is defined as utility which is "specific to the subject matter claimed," in contrast to "a general utility that would be applicable to the broad class of the invention." *M.P.E.P.* § 2107.01 I. Appellants submit that the evidence of differential expression of the PRO1753 gene and polypeptide in certain types of tumor cells, along with the declarations and references discussed above, provide a specific utility for the claimed polypeptides.

As discussed above, there are significant and reliable data which show that the gene for the PRO1753 polypeptide is expressed at least two-fold higher in esophageal tumor tissue compared to normal esophageal tissue. These data are strong evidence that the PRO1753 gene and polypeptide are associated with esophageal tumors. Thus, contrary to the assertions of the Examiner, Appellants have provided evidence associating the PRO1753 gene and polypeptide with a specific disease. The asserted utility for the claimed polypeptides as diagnostic tools for cancer, particularly esophageal cancer, is a specific utility – it is not a general utility that would apply to the broad class of polypeptides.

8. ***The Examiner's Response to Appellants' Evidence is Insufficient to Rebut Appellants' Arguments***

The Examiner has stated that the Grimaldi and Polakis declarations and other references cited by Appellants are insufficient to overcome the rejection of claims based on 35 U.S.C. §§ 101. *Final Office Action* at 5, 15, 17 and 18-19.

a. The Examiner's response to the First Grimaldi Declaration

Appellants have relied on the first Grimaldi Declaration to further explain that the semi-quantitative analysis employed to generate the data of Example 18 is sufficient to determine if a gene is over- or under-expressed in tumor cells compared to corresponding normal tissue. The declaration states that any visually detectable difference seen between two samples is indicative of at least a two-fold difference in cDNA between the tumor tissue and the counterpart normal tissue, and that the results of the gene expression studies indicate that the genes of interest "can be used to differentiate tumor from normal." *First Grimaldi Declaration* at ¶ 7.

The Examiner rejects the first Grimaldi Declaration, arguing that:

> The expression of all polynucleotides or polypeptides from a tumor sample or any other sample can invariably be classified as either increased, decreased, non-existent, or unchanged as compared to some standard level of expression. It can then be asserted that all proteins or polynucleotides that are expressed in this manner can be used to detect or characterize the tumor or other sample. Such utilities are analogous to the assertion that a particular protein can be employed as a molecular weight marker, which is neither a specific or substantial utility. *Final Office Action* at 6 (emphasis added).

This response fails to address the opinions expressed in the first Grimaldi Declaration. Neither the Appellants nor the first Grimaldi Declaration are asserting that "non-existent, or unchanged" expression provides utility for a gene or encoded polypeptide. This is made clear by the first Grimaldi declaration which states, "Example 18 reports the results of PCR analysis conducted as part of the investigating of several newly discovered DNA sequences.... The analyses were designed to determine whether a difference exists between gene expression in normal tissues as compared to tumor in the same tissue type. *First Grimaldi Declaration* at ¶ 4 (emphasis added). The declaration continues: "In differential gene expression studies, one looks

for genes whose expression levels differ significantly under different conditions, for example, in normal versus diseased tissue." *Id.* at ¶ 6 (emphasis added). Similarly, the declaration states that "the assay was designed to indicate whether a difference exists between normal tissue and tumor tissue of the same type. If a difference is detected, this indicates that the gene and its corresponding polypeptide and antibodies against the polypeptide are useful for diagnostic purposes, to screen samples to differentiate between normal and tumor." *Id.* at ¶ 7.

Appellants are asserting that because there is a difference in expression of the PRO1753 mRNA in esophageal tumor tissue compared to normal esophageal tissue, the PRO1753 nucleic acids have utility as diagnostic tools. In addition, as explained above, given the relationship between changes in mRNA level and changes in the level of the corresponding protein, one skilled in the art would expect PRO1753 polypeptide expression to be increased in esophageal tumor tissue compared to the corresponding normal tissue as well. Therefore the claimed polypeptides have utility as diagnostic tools.

This is not the same as an asserted utility of using a protein as a molecular weight marker since not all genes or proteins are differentially expressed in tumors. In addition, as detailed above, this is a specific utility since the PRO1753 gene and protein are differentially expressed in a specific type of tumor, namely esophageal tumors. Thus, the PTO's rejection of the first Grimaldi Declaration is based on a mischaracterization of Appellants' asserted basis for utility.

In rejecting the first Grimaldi Declaration, the Examiner also repeats his often made assertions that "no information is provided in the differential analysis of the PRO1753 polynucleotide expression regarding the level of expression, activity, or role in cancer of the PRO1753 polypeptide;" that "protein expression levels are not predictable from the mRNA expression levels;" and that "the skilled artisan would not know if or how PRO1753 polypeptide expression would change in cancer." *Id.*

Appellants remind that PTO that the evidence of the asserted utility can be indirect evidence which is reasonably correlated to the asserted utility, and that the asserted utility need not be established to an absolute or statistical certainty or even beyond a reasonable doubt – the standard is more likely than not true. *See M.P.E.P.* § 2107.02, part VII; *see also discussion infra* at § VII.B.9; *Fujikawa v. Wattanasin*, 93 F.3d 1559, 1565, 39 U.S.P.Q.2d 1895 (Fed. Cir. 1996)

("a 'rigorous correlation' need not be shown in order to establish practical utility; 'reasonable correlation' suffices"). As discussed above, Appellants have established that it is more likely than not that one of skill in the art would accept that changes in mRNA lead to corresponding changes in the encoded polypeptide. In light of all the evidence, one skilled in the art would find it more likely than not that the PRO1753 polypeptide is overexpressed in esophageal tumors based on the fact that the PRO1753 mRNA is overexpressed in esophageal tumors. Thus, contrary to the Examiner's assertions, Appellants have provided information regarding the expression of the PRO1753 polypeptide; changes in protein level are predictable from changes in mRNA levels; and one skilled in the art would know that PRO1753 polypeptide expression would be increased in esophageal cancer.

Finally, in rejecting the first Grimaldi Declaration, the Examiner states that "the examiner is not saying that Applicants must disclose the activity or role in cancer of the PRO1753 polypeptide. The examiner is saying that Applicants have not provided any information in the present specification regarding the level of expression, activity, or role in cancer of the PRO1753 polypeptide." *Id.* at 6-7 (emphasis added).

While the meaning of this apparently contradictory statement is unclear to Appellants, Appellants reiterate that a lack of a known activity or role for PRO1753 in cancer does not prevent its use as a diagnostic tool for cancer. The fact that there is no known translocation or mutation of PRO1753, for example, is irrelevant to whether its differential expression can be used to assist in diagnosis of cancer – one does not need to know why PRO1753 is differentially expressed, or what the consequence of the differential expression is, in order to exploit the differential expression to distinguish tumor from normal tissue.

The Revised Interim Utility Guidelines Training Materials available from the PTO's website establish that proteins which are differentially expressed in cancer have utility regardless of the lack of any known role in cancer. In Example 12 of the Guidelines, the hypothetical specification discloses a protein isolated from a cell membrane preparation which is the binding partner for protein X. Based on this relationship, the protein is designated receptor A. The function of X is not known, and the specification does not characterize the isolated protein with regard to its biological function or any disease. The Guidelines state that this disclosure is not

sufficient to establish utility. However, in a caveat, the facts are changed such that the specification also discloses that receptor A is expressed on melanoma cells but not on normal skin cells. The guidelines state that this additional disclosure is sufficient to establish utility, in spite of the fact that nothing is disclosed about the role of the protein in melanoma.

Finally, Appellants remind the PTO that "[o]ffice personnel must accept an opinion from a qualified expert that is based upon relevant facts whose accuracy is not being questioned." *M.P.E.P.* § 2107 (emphasis added). In addition, declarations relating to issues of fact should not be summarily dismissed as "opinions" without an adequate explanation of how the declaration fails to rebut the Examiner's position. *See In re Alton* 76 F.3d 1168 (Fed. Cir. 1996). The Examiner has not supplied any reasons or evidence to question the accuracy of the facts upon which Mr. Grimaldi based his opinion. Mr. Grimaldi has personal knowledge of the relevant facts, has based his opinion on those facts, and the PTO has offered no reason or evidence to reject either the underlying facts or his opinion.

In conclusion, Appellants submit that the evidence reported in Example 18, combined with the first Grimaldi Declaration, establish that there is at least a two-fold difference in PRO1753 cDNA between esophageal tumor tissue and normal esophageal tissue. Therefore, it follows that expression levels of the PRO1753 gene and polypeptide can be used to distinguish esophageal tumor tissue from normal esophageal tissue. The Examiner has not offered any significant arguments or evidence to the contrary.

b. The Examiner's response to the Second Grimaldi Declaration

To support their asserted utility, Appellants rely on paragraph 5 of the second Grimaldi Declaration: "Those who work in this field are well aware that in the vast majority of cases, when a gene is over-expressed...the gene product or polypeptide will also be over-expressed.... This same principal applies to gene under-expression." *Second Grimaldi Declaration* at ¶ 5. Further, Appellants have cited the statement that "increased mRNA expression is expected to result in increased polypeptide expression, and the detection of decreased mRNA expression is expected to result in decreased polypeptide expression." *Id.*

The Examiner asserts that the second Grimaldi Declaration is insufficient to overcome the rejection based upon a lack of utility. *Final Office Action* at 15. The Examiner repeats the same arguments made in rejecting the first Grimaldi Declaration: that the expression of all polynucleotides or polypeptides can invariably be classified as either increased, decreased, non-existent, or unchanged as compared to some standard level of expression; and that "the examiner is not saying that Applicants must disclose the activity or role in cancer of the PRO1753 polypeptide. The examiner is saying that Applicants have not provided any information in the present specification regarding the level of expression, activity, or role in cancer of the PRO1753 polypeptide." *Id.* at 15-16.

Appellants repeat that it is the difference in expression of the PRO1753 mRNA in esophageal tumors compared to normal esophageal tissue which provides utility. Appellants maintain that the data in Example 18 are significant and reliable, and that based on the evidence as a whole, one of skill in the art would accept that it is more likely than not that the PRO1753 polypeptide is also differentially expressed in esophageal tumors. Therefore, Appellants have provided information regarding the expression level of the PRO1753 polypeptide in cancer. As to any lack of evidence regarding the role or activity of the PRO1753 polypeptide in cancer, Appellants reiterate that one does not need to know what role or activity PRO1753 plays in cancer in order to use its differential expression to distinguish tumor tissue from normal tissue.

The Examiner also rejects the second Grimaldi Declaration by arguing that the current situation is unlike the references cited in paragraph four of the declaration. The Examiner argues that unlike the situation in Grimaldi *et al.* (Blood (1989) 73(8):2081-5) there is no evidence that the present situation involves the cloning of a chromosomal breakpoint, and unlike Meeker *et al.* (Blood (1990) 76(2):285-9) and Singleton *et al.* (Pathol. Annu. (1992) 27(1):165-90) where protein expression was associated with disease state, "the present specification does not provide any testing of the level of expression, activity, or role in cancer of the PRO1753 polypeptide." *Final Office Action* at 16.

As an initial matter, Appellants note that they have not relied on paragraph four of the second Grimaldi Declaration to support their asserted utility. Following a discussion of the Grimaldi *et al.*, Meeker *et al.* and Singleton *et al.* references, the declaration states that "[i]f the

chromosomal aberration results in the aberrant expression of mRNA and the corresponding gene product...then the gene product is a promising target for cancer therapy, for example, by the therapeutic antibody approach." *Second Grimaldi Declaration* at ¶ 4. Thus, paragraph four and the cited references relate to a discussion of the development of targets for cancer therapy, not cancer diagnostics.

In addition, as discussed above, the fact that there is no known translocation or mutation of PRO1753 is irrelevant to whether its differential expression can be used to assist in diagnosis of cancer – one does not need to know why PRO1753 is differentially expressed, or what the consequence of the differential expression is, in order to exploit the differential expression to distinguish tumor from normal tissue. As to the Examiner's argument that unlike Meeker *et al.* and Singleton *et al.* the present specification does not provide any testing of the level of expression, activity, or role in cancer of the PRO1753 polypeptide, Appellants have provided information regarding changes in the expression of PRO1753 mRNA in cancer, and one of skill in the art would recognize that this is reasonably correlated to changes in PRO1753 protein expression such that Appellants have provided information regarding the expression of PRO1753 polypeptide in cancer.

Appellants remind the PTO that "[o]ffice personnel must accept an opinion from a qualified expert that is based upon relevant facts whose accuracy is not being questioned." *M.P.E.P.* § 2107 (emphasis added). In addition, declarations relating to issues of fact should not be summarily dismissed as "opinions" without an adequate explanation of how the declaration fails to rebut the Examiner's position. *See In re Alton* 76 F.3d 1168 (Fed. Cir. 1996). The Examiner has not supplied any reasons or evidence to reject either the underlying facts or opinion expressed in the second Grimaldi Declaration.

In conclusion, Appellants submit that the second Grimaldi declaration supports Appellants' assertion that one of skill in the art would be more likely than not to believe that the expression levels of the PRO1753 gene and polypeptide can be used to distinguish esophageal tumor tissue from normal esophageal tissue. The Examiner has not offered any relevant arguments or evidence to the contrary, and therefore has not provided a basis to reject the second Grimaldi Declaration.

c. The Examiner's response to the Polakis Declaration

Appellants have relied on the statement of paragraph six of the Polakis Declaration to support their asserted utility: "an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell. In fact, it remains a central dogma in molecular biology that increased mRNA levels are predictive of corresponding increased levels of the encoded protein." *Polakis Declaration* at ¶ 6.

The Examiner has rejected the Polakis Declaration as insufficient to overcome the utility rejection because "[t]he present application provides no information regarding the level of expression, activity, or role in cancer of the PRO1753 polypeptide. Only mRNA expression data was presented." *Final Office Action* at 17. The Examiner states that he has cited countervailing evidence to show that the significance or relevance of the disclosed PRO1753 mRNA expression in relation to cancer diagnosis or treatment is unknown, and secondly, that mRNA levels are not always consistent with protein levels. *Id.* at 17-18.

Appellants submit that these arguments do not address the opinions expressed in the Polakis Declaration. As the second Grimaldi Declaration, the Polakis Declaration, and the other supporting references cited by the Appellants make clear, information about changes in mRNA levels is sufficient for one of skill in the art to conclude that it is more likely than not that the protein levels are similarly changed. Therefore, Appellants have provided information regarding the level of expression of the PRO1753 polypeptide in cancer.

Second, as discussed above, the Examiner's "countervailing evidence" does not support his arguments – Hu and Wang do not relate to using differentially expressed genes as molecular markers; Haynes and Gygi examined steady-state mRNA and protein levels across different genes, not the relationship between changes in mRNA level and changes in protein levels; and Hancock offers no support for the Examiner's position even if it is consistent with Haynes. Therefore, the Examiner has not cited any relevant "countervailing evidence" on which to rely in rejecting the conclusions of the Polakis Declaration.

The Examiner also rejects the Polakis Declaration, stating that a "dogma" is an authoritative principle, belief, or statement of ideas or opinion, especially one considered to be absolutely true. *Final Office Action* at 18. The Examiner asserts that Haynes and Hancock provide evidence that Dr. Polakis' asserted dogma is not "absolutely true" and that the skilled artisan would have a legitimate basis to doubt the utility of the antibodies which bind to the PRO1753 polypeptide based solely on the differential analysis of PRO1753 mRNA expression. *Id.*

First, Appellants have emphasized that they do not need to prove an exact or absolute correlation between changes in mRNA and changes in protein levels to establish their asserted utility – the correlation does not need to be "absolutely true":

> [T]he applicant does not have to provide evidence sufficient to establish that an asserted utility is true "beyond a reasonable doubt." Nor must the applicant provide evidence such that it establishes an asserted utility as a matter of statistical certainty. Instead, evidence will be sufficient if, considered as a whole, it leads a person of ordinary skill in the art to conclude that the asserted utility is more likely than not true. *M.P.E.P.* § 2107.02, part VII (emphasis in original, citations omitted).

In response to similar arguments made by Appellants, the Examiner stated that the rejection of Appellants' asserted utility "is not based upon 'an absolute certainty,' 'statistical certainty,' 'absolute predictability,' 'an invariable exact correlation,' or a universal correlation." *Final Office Action* at 3.

In spite of this apparent acknowledgement of the proper standard for establishing utility, the Examiner rejects the Polakis declaration, arguing that Haynes and Hancock allegedly prove that Dr. Polakis' asserted dogma is not "absolutely true." Further, the Examiner states throughout the Office Action that the claimed polypeptides lack utility because the Examiner has provided evidence showing "that protein levels are not *always* consistent with mRNA levels." *Final Office Action* at 3, 4, 6, 9, 10, 18 and 23 (emphasis added.) Thus, it is clear that the Examiner is continuing to require that Appellants establish their asserted utility as a matter of absolute or statistical certainty.

This standard is inconsistent with the Utility Guidelines and governing case law. Therefore, even if the Polakis "dogma" is not "absolutely true," this is not a basis to reject the

asserted utility. Additionally, as noted above, the references relied on by the Examiner (Haynes and Hancock) do not actually provide any evidence that is contrary to the conclusions of the Polakis Declaration.

Appellants remind the PTO that "[o]ffice personnel must accept an opinion from a qualified expert that is based upon relevant facts whose accuracy is not being questioned." *M.P.E.P.* § 2107 (emphasis added). In addition, declarations relating to issues of fact should not be summarily dismissed as "opinions" without an adequate explanation of how the declaration fails to rebut the Examiner's position. *See In re Alton* 76 F.3d 1168 (Fed. Cir. 1996). The Examiner has not supplied any reasons or evidence to reject either the underlying facts or opinion expressed in the Polakis Declaration.

In conclusion, Appellants submit that the Polakis Declaration supports Appellants' assertion that one of skill in the art would be more likely than not to believe that the expression levels of the PRO1753 gene and polypeptide can be used to distinguish esophageal tumor tissue from normal esophageal tissue. The Examiner has not offered any relevant arguments or evidence to the contrary, and therefore has not provided a basis to reject the Polakis Declaration.

d. The Examiner's response to the Molecular Biology of the Cell and Lewin references submitted by Appellants

Appellants submitted the Molecular Biology of the Cell and Lewin references in support of their assertion that, in general, changes in mRNA expression correlates with changes in expression of the encoded polypeptide. Although the Examiner acknowledges the teachings in the Molecular Biology of the Cell and Lewin, he notes that Molecular Biology of the Cells states that "other controls can act later in the pathway from DNA to protein to modulate the amount of gene product that is made" and Lewin states that "production of RNA cannot inevitably be equated with production of protein." *Final Office Action* at 18-19. According to the Examiner, both Molecular Biology of the Cell and Lewin support and are consistent with the Examiner's position that that the skilled artisan would not know if or how PRO1753 polypeptide expression would change in cancer. *Id.* at 19.

Appellants have acknowledged that changes in mRNA do not *always* result in changes in protein level. However, Appellants do not have to prove utility to an "absolute certainty" or "beyond a reasonable doubt." Instead, one of skill in the art must believe that the asserted utility is more likely than not true. The Examiner fails to address the teachings of Molecular Biology of the Cell which provides that "[f]or most genes transcriptional controls are paramount," *Cell 3rd* at 403 and *Cell 4th* at 379 (emphasis added); that "the initiation of gene transcription are the predominant form of regulation for most genes," *Cell 3rd* at 453; that "a cell can change (or regulate) the expression of each of its genes according to the needs of the moment – most obviously by controlling the production of its mRNA," *Cell 4th* at 302 (emphasis added). Similarly, Figure 6-90 on page 364 of Cell 4th illustrates the path from gene to protein. Similarly, the Examiner has not addressed the teachings of Lewin, which states "having acknowledged that control of gene expression can occur at multiple stages, and that production of RNA cannot inevitably be equated with production of protein, it is clear that the overwhelming majority of regulatory events occur at the initiation of transcription." *Genes VI* at 847-848 (emphasis added).

The PTO does not provide a basis for failing to address evidence of the "overwhelming majority" and "paramount" controls of gene expression, and instead focuses on "other controls" that can act later. Appellants submit that a full consideration of the submitted evidence would lead to a conclusion that the "overwhelming majority" and "paramount" controls have a greater influence on gene expression than "other controls" that can act later. Accordingly, full consideration of this evidence would lead to a conclusion that mRNA levels are paramount in influencing polypeptide expression levels, and that one of skill in the art would conclude that it is more likely than not that changes in mRNA lead to changes in the corresponding protein. Therefore, the Cell and Lewin references support Appellants' asserted utility, and the Examiner has not provided any basis to reject Appellants' conclusions based on these references.

e. The Examiner's response to the Zhigang reference submitted by Appellants

Appellants submitted the Zhigang reference in support of their position that, in general, differential mRNA expression correlates with differential expression of the encoded polypeptide.

The Examiner acknowledges that Zhigang presents data showing a high degree of correlation between changes in PSCA mRNA and protein expression. However, the Examiner asserts that exceptions were noted. *Final Office Action* at 19. Thus, according to the Examiner, Zhigang supports and is consistent with the Examiner's position that "the skilled artisan would not know if or how PRO1753 polypeptide expression would change in cancer." *Id.*

Zhigang reported that the correlation between mRNA expression and protein expression occurred in 93% of the samples tested. Appellants submit that there is no requirement to provide evidence sufficient to establish an asserted utility as a matter of statistical certainty. "Instead, evidence will be sufficient if, considered as a whole, it leads a person of ordinary skill in the art to conclude that the asserted utility is more likely than not true." *M.P.E.P.* at § 2107.02, part VII (2004) (emphasis in original, internal citations omitted).

In previous responses, Appellants have emphasized that they do not need to prove an exact or absolute correlation between changes in mRNA and changes in protein levels to establish their asserted utility. In response, the Examiner stated that the final rejection "is not based upon 'an absolute certainty,' 'statistical certainty,' 'absolute predictability,' 'an invariable exact correlation,' or a universal correlation." *Final Office Action* at 3.

In spite of this apparent acknowledgement of the proper standard for establishing utility, the Examiner holds that a showing of 93% correlation between mRNA and polypeptide levels is an indication of support for unpredictability of the relationship between changes in mRNA and polypeptide levels. As noted above, the Examiner states throughout the Office Action that the claimed polypeptides lack utility because the Examiner has provided evidence showing "that protein levels are not always consistent with mRNA levels." *Final Office Action* at 3, 4, 6, 9, 10, 18 and 23 (emphasis added.) Thus, it is clear from the Examiner's evaluation of the evidence of record that the Examiner is requiring that Appellants establish their asserted utility as a matter of statistical or absolute certainty. This standard is inconsistent with the Utility Guidelines and governing case law. Since the Examiner's rejection of the teachings of the Zhigang reference is based on this incorrect standard, the Examiner has not provided any basis to reject Appellants' conclusion that Zhigang supports Appellants' asserted utility.

f. ***The Examiner's response to the Meric reference submitted by Appellants***

Appellants submitted the Meric reference in support of their position that, in general, changes in mRNA expression correlate with changes in expression of the encoded polypeptide. Appellants rely on the statement in Meric that "[t]he fundamental principle of molecular therapeutics in cancer is to exploit the differences in gene expression between cancer cells and normal cells." *Meric* at 971 (emphasis added).

While the Examiner acknowledges this statement, the Examiner asserts that the specification does not provide any testing of the level of expression, activity, or role in cancer of the PRO1753 polypeptide. *Final Office Action* at 19. Therefore, according to the Examiner, the differences in PRO1753 polypeptide expression between cancer cells and normal cells are unknown, and thus they are not exploitable. *Id.* The Examiner states that Meric also acknowledges that several alterations in translational control occur in cancer and that gene expression is quite complicated. *Id.* at 19-20. Thus, the Examiner asserts that Meric suggests that protein levels can be modulated independently of the level of mRNA, and therefore Meric supports and is consistent with his position that "the skilled artisan would not know if or how PRO1753 polypeptide expression would change in cancer." *Id.* at 20.

Meric does teach that mutations of genes as well as alternate splicing and alternate transcription start sites can lead to altered translation efficiency in certain cancer cells. *Meric* at 973-974. As support, Meric cites three examples of point mutations, and four examples of alternate splicing. *Id.* at 974. However, the Examiner has not shown, and there is no evidence, that the PRO1753 mRNA is either mutated, alternately spliced, or has an alternate transcription start site. Nor has the Examiner established that point mutations or alternate splice variants leading to changes in translation efficiency are common in cancer, or common in esophageal tumors in particular. These few examples are not sufficient to provide evidence that one skilled in the art would reasonably doubt Appellants' asserted utility, or reject the teaching of Appellants' supporting references and declarations.

As the supporting references and declarations Appellants have discussed above make clear, regulation of mRNA levels is the predominant control mechanism for the majority of

genes. Meric supports this assertion because "[t]he **fundamental principle** of molecular therapeutics in cancer is to exploit the differences in gene expression between cancer cells and normal cells." *Meric* at 971 (emphasis added). The only reason differences in mRNA expression between cancer and normal tissue is of any interest in studying the mechanism of cancer formation and growth is because mRNA encodes protein. If there were no general correlation between differences in mRNA and differences in protein, there would be no reason to study changes in mRNA. Thus, Meric supports Appellants' assertion that those of skill in the art recognize that there is a general correlation between changes in mRNA levels and changes in the level of the corresponding protein.

As the Examiner has not offered any significant reasoning or evidence to the contrary, the teachings of Meric and Appellants' conclusion that it supports their asserted utility should be accepted as true.

g. Conclusion

In conclusion, Appellants have offered sufficient evidence to establish that it is more likely than not that one of skill in the art would believe that because the PRO1753 mRNA is overexpressed in esophageal tumors compared to normal esophageal tissue, the PRO1753 polypeptide will also be overexpressed in esophageal tumors compared to normal esophageal tissue. This differential expression of the PRO1753 polypeptide makes it useful as a diagnostic tool for cancer. The declarations and references cited by Appellants clearly support their asserted utility, and the Examiner has offered no relevant arguments or evidence to the contrary. In short, none of the Examiner's responses to Appellants' supporting evidence are sufficient to rebut Appellants' asserted utility.

9. *The Courts have held that the Utility Requirement was Satisfied in Similar Cases*

The seminal decision interpreting the utility requirement of 35 U.S.C. § 101 is *Brenner v. Manson*, 383 U.S. 519, 148 U.S.P.Q. 689 (1966). At issue in *Brenner* was a claim to "a chemical process which yields an already known product whose utility – other than as a possible object of

scientific inquiry – ha[d] not yet been evidenced." *Id.* at 529, 148 U.S.P.Q. at 693. The Patent Office rejected the claimed process for lack of utility because the product produced by the claimed process had no known use. *See id.* at 521-22, 148 U.S.P.Q. at 690. On appeal, the Court of Customs and Patent Appeals reversed, holding "where a claimed process produces a known product it is not necessary to show utility for the product." *Id.* at 522, 148 U.S.P.Q. at 691.

In reviewing the lower court's decision, the Court made its oft quoted statement that "[t]he basic quid pro quo contemplated by the Constitution and the Congress for granting a patent monopoly is the benefit derived by the public from an invention with substantial utility. Unless and until a process is refined and developed to this point – where specific benefit exists in currently available form – there is insufficient justification for permitting an Appellant to engross what may prove to be a broad field." *Id.* at 534-35, 148 U.S.P.Q. at 695.

The first opinion of the C.C.P.A. applying *Brenner* was *In re Kirk*, 376 F.2d 936, 153 U.S.P.Q. 48 (C.C.P.A. 1967). The invention claimed in *Kirk* was a set of steroid derivatives said to have valuable biological properties and to be of value "in the furtherance of steroidal research and in the application of steroidal materials to veterinary or medical practice." *Id.* at 938, 153 U.S.P.Q. at 50. In affirming the claim rejection based on a lack of utility, the court held that the "nebulous expressions 'biological activity' or 'biological properties'" did not adequately convey how to use the claimed compounds." *Id.* at 941, 153 U.S.P.Q. at 52. The court also rejected Appellants' supporting affidavit, stating, "the sum and substance of the affidavit appears to be that one of ordinary skill in the art would know 'how to use' the compounds to find out in the first instance whether the compounds are – or are not – in fact useful or possess useful properties, and to ascertain what those properties are." *Id.* at 942, 153 U.S.P.Q. at 53.

Since these early decisions, the courts have continued to clarify what is sufficient to satisfy the utility requirement. Three more recent decisions are of particular relevance to the instant application: *Nelson v. Bowler*, 626 F.2d 853, 206 U.S.P.Q. 881 (C.C.P.A. 1980), *Cross v. Iizuka*, 753 F.2d 1040, 224 U.S.P.Q. 739 (Fed. Cir. 1985), and *Fujikawa v. Wattanasin*, 93 F.3d 1559, 39 U.S.P.Q. 2d 1895 (Fed. Cir. 1996).

The earliest of these cases, *Nelson v. Bowler*, involved an interference between two applications related to derivatives of naturally occurring prostaglandins (PG). *Nelson*, 626 F.2d

at 854-55. The issue was whether Nelson had shown at least one utility for the compounds at issue to establish an actual reduction to practice. *Id.* at 855. The Appellants relied on two tests to prove practical utility: an *in vivo* rat blood pressure (BP) test and an *in vitro* gerbil colon smooth muscle stimulation (GC-SMS) test. In the BP test, the blood pressure of anesthetized rats was recorded on a polygraph chart to determine whether an injected compound had any effect. Responses were categorized as either a depressor (lowering) effect or a pressor (elevating) effect. *Id.* In the GC-SMS test a section of colon was excised from a freshly-killed gerbil for suspension in a physiological solution, and a lever arm was connected to the colon in such a way that any contraction was recorded as a polygraph trace. *Id.* The Board held that Nelson had not shown adequate proof of practical utility, characterizing the tests as "rough screens, uncorrelated with actual utility." *Id.* at 856.

On appeal the C.C.P.A. reversed, holding that the Board "erred in not recognizing that tests evidencing pharmacological activity may manifest a practical utility even though they may not establish a specific therapeutic use." *Id.* The Court stated that "practical utility" was characterized as a use of the claimed discovery in a manner which provides some immediate benefit to the public, establishing the following rule:

> Knowledge of the pharmacological activity of any compound is obviously beneficial to the public. It is inherently faster and easier to combat illnesses and alleviate symptoms when the medical profession is armed with an arsenal of chemicals having known pharmacological activities. Since it is crucial to provide researchers with an incentive to disclose pharmacological activities in as many compounds as possible, we conclude that adequate proof of <u>any</u> such activity constitutes a showing of practical utility. *Id.* (emphasis added).

The Court rejected Bowler's argument that the BP and GC-SMS tests are inconclusive showings of pharmacological activity since confirmation by statistically significant means did not occur until after the critical date. The Court stated that "a rigorous correlation is <u>not necessary</u> where the test for pharmacological activity is <u>reasonably indicative</u> of the desired response." *Id.* (emphasis added). The Court concluded that a "<u>reasonable correlation</u>" between the observed properties and the suggested use was sufficient to establish practical utility. *Id.* at 857 (emphasis added).

The sufficiency of a "reasonable correlation" in establishing utility was affirmed by the Court of Appeals for the Federal Circuit in *Cross v. Iizuka*, 753 F.2d 1040, 224 U.S.P.Q. 739 (Fed. Cir. 1985). In *Cross*, the subject of the interference before the Court was imidazole derivative compounds which inhibit the synthesis of thromboxane synthetase, an enzyme which leads to the formation of thromboxane A_2. At the time the applications were filed, thromboxane A_2 was postulated to be involved in platelet aggregation, which was associated with several deleterious conditions. *Id.* at 1042.

The question before the Board and reviewed by the Court was whether Iizuka was entitled to the benefit of his Japanese priority application. *Id.* The Japanese application disclosed that the imidazole derivatives showed strong inhibitory action for thromboxane synthetase from human or bovine platelet microsomes, an *in vitro* utility. *Id.* at 1043. Relying in part on *Nelson,* the Board held that tests evidencing pharmacological activity may manifest a practical utility even though they may not establish a specific therapeutic use, and concluded that the *in vitro* tests were sufficient to establish a practical utility. *Id.*

On appeal, Cross argued that the basic *in vitro* tests conducted in cellular fractions did not establish a practical utility for the claimed compounds, and that more sophisticated *in vitro* or *in vivo* tests were necessary to establish a practical utility. *Id.* at 1050. The Court rejected this argument, noting that adequate proof of any pharmaceutical activity constitutes a showing of practical utility. *Id.* The Court accepted the argument that initial testing of compounds is widely done *in vitro*:

> [*I*]*n vitro* results...are generally predictive of *in vivo* test results, i.e., there is a reasonable correlation therebetween. Were this not so, the testing procedures of the pharmaceutical industry would not be as they are. Iizuka has not urged, and rightly so, that there is an invariable exact correlation between *in vitro* test results and *in vivo* test results. Rather, Iizuka's position is that successful *in vitro* testing for a particular pharmacological activity establishes a significant probability that *in vivo* testing for this particular pharmacological activity will be successful. *Id.* (emphasis added).

The Court also noted that in previous decisions, its predecessor court had accepted evidence of *in vivo* utility as sufficient to establish practical utility. The Court reasoned that:

> This *in vivo* testing is but an intermediate link in a screening chain which may eventually lead to the use of the drug as a therapeutic agent in humans. We

> perceive no insurmountable difficulty, under appropriate circumstances, in finding that the first link in the screening chain, *in vitro* testing, may establish a practical utility for the compound in question. Successful *in vitro* testing will marshal resources and direct the expenditure of effort to further *in vivo* testing of the most potent compounds, thereby providing an immediate benefit to the public, analogous to the benefit provided by the showing of an *in vivo* utility. *Id.* at 1051, *citing Nelson,* 626 F.2d at 856 (emphasis added).

Based on this reasoning, the Court affirmed the decision of the Board, stating that "based upon the relevant evidence as a whole, there is a reasonable correlation between the disclosed *in vitro* utility and an *in vivo* activity, and therefore a rigorous correlation is not necessary where the disclosure of pharmacological activity is reasonable based upon the probative evidence." *Id.* at 1050 (emphasis added). The Court therefore held that the disclosed *in vitro* utility was "sufficient to comply with the practical utility requirement of § 101." *Id.* at 1051.

The holdings of *Nelson* and *Cross* were more recently affirmed in *Fujikawa v. Wattanasin*, 93 F.3d 1559, 39 U.S.P.Q.2d 1895 (Fed. Cir. 1996). In *Fujikawa*, the Court again affirmed the notion that initial screens of compounds provide a practical utility even though they may not provide a therapeutic use because "'[i]t is inherently faster and easier to combat illnesses and alleviate symptoms when the medical profession is armed with an arsenal of chemicals having known pharmacological activities.'" *Id.* at 1564, *quoting Nelson,* 626 F.2d at 856. The Court noted that it may be difficult to predict whether novel compounds will exhibit pharmacological activity, and consequently testing is often required to establish practical utility. *Id.* However the Court went on to state:

> But the test results need not absolutely prove that the compound is pharmacologically active. All that is required is that the tests be "*reasonably* indicative of the desired [pharmacological] response." In other words, there must be a sufficient correlation between the tests and an asserted pharmacological activity so as to convince those skilled in the art, to a reasonable probability, that the novel compound will exhibit the asserted pharmacological behavior." *Id.* (internal citations omitted, underline emphasis added, italics in original).

On appeal, Fujikawa argued that Wattanasin had failed to establish an adequate correlation between the *in vitro* and *in vivo* results to permit Wattanasin to rely on positive *in vitro* results to establish a practical utility. The Court stated that the Board relied on testimony from those skilled in the art that the *in vitro* results convinced the experts that the claimed

compounds would exhibit the desired pharmacological activity when administered *in vivo*, including testimony that *in vivo* activity is typically highly correlatable to a compound's *in vitro* activity in the field. *Id.* at 1565. To overcome this evidence and counter the Board's decision, Fujikawa pointed to the testimony of its expert that "there is a reasonable element of doubt that some elements may be encountered which are active in the *in vitro* assay, but yet inactive in the *in vivo* assay." *Id.*

The Court rejected this argument: "Of course, it is possible that some compounds active *in vitro* may not be active *in vivo.* But, as our predecessor court in *Nelson* explained, a 'rigorous correlation' need not be shown in order to establish practical utility; 'reasonable correlation' suffices." *Id.* (emphasis added). The Court also rejected Fujikawa's reliance on two articles. The Court noted that while one article taught that "*in vitro* testing is sometimes not a good indicator of how potent a compound will be *in vivo,* it does imply that compounds which are active *in vitro* will normally exhibit some *in vivo* activity." *Id.* at 1566. Similarly, the Court noted that the second article expressly stated that "[f]or most substances, although not for all, the relative potency determined in *in vitro* ... parallels the *in vivo* activity." *Id.*

The Court concluded that the facts in the case were analogous to the ones in *Cross* where the court relied on a known reasonable correlation between *in vitro* tests and *in vivo* activity, and therefore affirmed the Board's decision that Wattanasin had established a practical utility with the *in vitro* results. *Id.* at 1565-66.

The *Nelson*, *Cross*, and *Fujikawa* cases are very similar to the present case. The reasoning of the courts in all three cases that "'[i]t is inherently faster and easier to combat illnesses and alleviate symptoms when the medical profession is armed with an arsenal of chemicals having known pharmacological activities'" applies to the asserted utility for the claimed polypeptides. *Fujikawa*, 93 F.3d at 1564, *quoting Nelson,* 626 F.2d at 856; *see also Cross*, 753 F.2d at 1051 ("Successful *in vitro* testing will marshal resources and direct the expenditure of effort to further *in vivo* testing of the most potent compounds, thereby providing an immediate benefit to the public, analogous to the benefit provided by the showing of an *in vivo* utility."). Like pharmaceutical compounds, nucleic acids, polypeptides, and antibodies which are associated with cancer will make it inherently faster and easier to combat cancer. The

greater the number of biological markers of cancer medical professionals have access to, the more accurate and detailed a diagnosis they can make. The determination that a gene is differentially expressed in cancer constitutes at least as significant a development in the field of cancer diagnostics as *in vitro* screening for pharmaceutical activity. *See Cross*, 753 F.2d at 1051 ("the first link in the screening chain, *in vitro* testing, may establish a practical utility for the compound in question. Successful *in vitro* testing will marshal resources and direct the expenditure of effort to further *in vivo* testing of the most potent compounds, thereby providing an immediate benefit to the public.").

In addition, like *in vitro* tests in the pharmaceutical industry, those of skill in the field of biotechnology rely on the reasonable correlation that exists between changes in gene expression and changes in protein expression (see discussion supra). Were there no reasonable correlation between the two, the techniques that measure gene levels such as microarray analysis, differential display, and quantitative PCR would not be so widely used by those in the art. *See Second Grimaldi Declaration* at ¶ 5. As in *Cross*, Appellants here do not argue that there is "an invariable exact correlation" between gene expression and protein expression. *See Cross*, 753 F.2d at 1050. Instead, Appellants' position detailed above is that a measured change in gene expression in cancer cells establishes a "significant probability" that the expression of the encoded polypeptide in cancer will also be changed based on "a reasonable correlation therebetween." *Id.*; *see also Fujikawa*, 93 F.3d at 1565 ("a 'rigorous correlation' need not be shown in order to establish practical utility; 'reasonable correlation' suffices"); *Nelson*, 626 F.2d at 857 (holding that "a rigorous correlation is not necessary" and that a "reasonable correlation" will suffice).

Also of importance is the Court's rejection of the notion that any *in vitro* testing must be statistically significant to support a practical utility. *Nelson*, 626 F.2d at 857. Likewise, qualitative characterizations of a test compound as either increasing or decreasing blood pressure was acceptable. *Id.* at 855 (stating that responses were categorized as either a depressor (lowering) effect or a pressor (elevating) effect). This is similar to the data in Example 18, where the change in mRNA levels is described as "more highly expressed."

There are additional similarities. In *Fujikawa*, the Board and Court rejected the argument that there was no utility because there was no exact correlation between the *in vitro* and *in vivo* results in spite of supporting testimony and references. *Fujikawa*, 93 F.3d at 1565-66. Like the two references rejected by the Board and Court in *Fujikawa*, portions of the references of record may suggest that the correlation between changes in mRNA levels and protein levels is not exact. But like the references in *Fujikawa*, the majority of each reference supports Appellants' assertion, and Appellants have submitted the declaration of two experts in the field which state that those skilled in the art rely on the correlation between changes in mRNA and protein. *See Second Grimaldi Declaration* at ¶ 5; *Polakis Declaration* at ¶ 6. Thus, as was the case in *Fujikawa*, although there may be some evidence that the correlation relied on is not exact, the declarations and numerous references submitted by Appellants are more than enough evidence to establish that there is a "reasonable correlation" between changes in mRNA levels and changes in the level of the encoded protein.

In conclusion, Appellants have asserted that the claimed polypeptides are useful for the diagnosis of cancer, particularly esophageal cancer, based on the data in Example 18. This utility is far beyond the nebulous expressions "biological activity" or "biological properties" rejected in *In re Kirk*, 376 F.2d 936, 153 U.S.P.Q. 48 (C.C.P.A. 1967). Like *Nelson*, *Cross*, and *Fujikawa*, Appellants have asserted a utility which relies on a reasonable correlation between the data disclosed in the application and the asserted utility. The fact that there may be limited evidence that the correlation is not exact does not invalidate Appellants' showing of utility since the correlation need not be a rigorous or exact one. Considering the relevant evidence as a whole, Appellants have provided sufficient evidence to establish a reasonable correlation between changes in the level of mRNA and corresponding changes in the level of the encoded polypeptide. Therefore the claimed polypeptides have a practical utility as diagnostic tools for esophageal cancer.

10. *Utility – Conclusion*

Appellants' asserted utility for the claimed polypeptides as diagnostic tools for cancer corresponds in scope to the subject matter sought to be patented and therefore "must be taken as

sufficient to satisfy the utility requirement of § 101 for the entire claimed subject." *In re Langer*, 503 F.2d 1380, 1391, 183 U.S.P.Q. 288, 297 (C.C.P.A. 1974). The Examiner's unsupported arguments and references are not sufficient evidence to make a *prima facie* showing that "one of ordinary skill in the art would reasonably doubt the asserted utility." *In re Brana*, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995).

And even if the Examiner has established a *prima facie* case, Appellants have offered sufficient rebuttal evidence in the form of expert declarations and references, which, when considered as a whole, establish that it is more likely than not that the asserted utility is true. *See In re Oetiker*, 977 F.2d 1443, 1445, 24 U.S.P.Q.2d 1443, 1444 (Fed. Cir. 1992) (stating that the evidentiary standard to be used throughout *ex parte* examination in setting forth a rejection is a preponderance of the evidence, or "more likely than not" standard); *M.P.E.P.* at § 2107.02, part VII ("evidence will be sufficient if, considered as a whole, it leads a person of ordinary skill in the art to conclude that the asserted utility is more likely than not true.") (emphasis in original).

Finally, the courts' decisions in similar cases make clear that the evidence provided by Appellants is sufficient to establish the asserted utility. The evidence does not need to be direct evidence, nor does it need to provide an exact correlation between the submitted evidence and the asserted utility. Instead, evidence which is "reasonably" correlated with the asserted utility is sufficient. *See Fujikawa*, 93 F.3d at 1565 ("a 'rigorous correlation' need not be shown in order to establish practical utility; 'reasonable correlation' suffices"); *Cross*, 753 F.2d at 1050 (same); *Nelson*, 626 F.2d at 857 (same). Considering the evidence as a whole in light of the relevant case law, the Board should find that Appellants have established at least one specific, substantial, and credible utility, and the Examiner's rejection of the pending claims as lacking utility should be reversed.

C. Enablement Rejection – Detailed Argument

The second issue before the Board is whether Appellants have enabled the pending claims such that one of skill in the art would be able to make and use the claimed invention. The Examiner has rejected pending Claims 6-8 and 11-17 under 35 U.S.C. §112, first paragraph, arguing that because the claimed invention is not supported by either a specific or substantial

asserted utility or a well-established utility, one skilled in the art clearly would not know how to use the claimed invention. *See First Office Action* at 24.

In addition, the Examiner has separately rejected pending Claims 12-17 under 35 U.S.C. 112, first paragraph, on the assertion that they contain subject matter which was not described in the specification in such a way as to enable one skilled in the art to make and/or use the invention. *Office Action* at 25.

As an initial matter, for the reasons discussed above, Appellants maintain that the claimed polypeptides possess a specific, substantial and credible utility. Accordingly, Appellants maintain that one of skill in the art would know how to use the claimed invention, and to the extent that the enablement rejection is based on a lack of utility, it should be reversed.

In response to the separate enablement rejection of Claims 12-17, Appellants submit that Claims 12-17 are enabled such that one of skill in the art could make and use the claimed polypeptides without undue experimentation. With respect to Claims 12-13, how to make chimeric polypeptides comprising the polypeptide of SEQ ID NO:110 and the polypeptide encoded by the cDNA deposited under ATCC accession number 203535 is within the skill in the art. Similarly, with respect to Claims 14-17, it is well within the skill of those in the art to make polypeptides that are at least 95% identical to SEQ ID NO:110 and the polypeptide encoded by ATCC 203535, and it is well within those of skill in the art to make antibodies which are specific to a disclosed sequence. *See In re Wands*, 858 F.2d 731 (reversing the Board's decision of non-enablement and holding that as of 1980, undue experimentation was not required to make high-affinity monoclonal antibodies to a target peptide). Thus, one of skill in the art would be able to make the claimed polypeptides without undue experimentation.

As described above, Appellants assert that the claimed polypeptides are useful as diagnostic tools for cancer, particularly esophageal cancer. This use is based on the disclosure in Example 18 of the instant application that the nucleic acid encoding the PRO1753 polypeptide is expressed at least two-fold higher in esophageal tumor compared to normal esophageal tissue. As detailed above, it is well-established that changes in expression levels of mRNA leads to corresponding changes in expression levels of the encoded polypeptide, and thus it is likely that the PRO1753 polypeptide also is differentially expressed in esophageal tumors. Thus, based on

the disclosure in the application, one of skill in the art would be able to use the claimed polypeptides as diagnostic tools, for example to make antibodies to distinguish suspected esophageal tumor tissue from normal esophageal tissue without undue experimentation.

1. ***Enablement – Legal Standard***

An application enables the claims "if one skilled in the art, after reading the[] disclosure[], could practice the invention claimed ... without undue experimentation." *Chiron Corp. v. Genentech, Inc.*, 363 F.3d 1247, 1253 (Fed. Cir. 2004). "But the question of undue experimentation is a matter of degree. The fact that some experimentation is necessary does not preclude enablement; what is required is that the amount of experimentation 'must not be unduly extensive.'" *PPG Indus., Inc. v. Guardian Indus., Corp.*, 75 F.3d 1558, 1564 (Fed. Cir. 1996) (quoting *Atlas Powder Co. v. E.I. DuPont de Nemours & Co.*, 750 F.2d 1569, 1576 (Fed. Cir. 1984)).

While the application must enable one of ordinary skill in the art to practice the full scope of the claimed invention, "[t]hat is not to say that the specification itself must necessarily describe how to make and use every possible variant of the claimed invention, for the artisan's knowledge of the prior art and routine experimentation can often fill gaps, interpolate between embodiments, and perhaps even extrapolate beyond the disclosed embodiments, depending upon the predictability of the art." *AK Steel Corp. v. Sollac*, 344 F.3d 1234, 1244 (Fed. Cir. 2003).

"Enablement is not precluded by the necessity for some experimentation such as routine screening. However, experimentation needed to practice the invention must not be undue experimentation. The key work is 'undue,' not 'experimentation.'" *In re Wands* 858 F.2d 731, 736-7, 8 U.S.P.Q.2d 1400, (Fed. Cir. 1988), citations omitted.

> It is equally clear that a rejection based on "lack of utility," whether grounded upon 35 U.S.C. 101 or 35 U.S.C. 112, first paragraph, rests on the same basis (i.e., the asserted utility is not credible). To avoid confusion, any rejection that is imposed on the basis of 35 U.S.C. 101 should be accompanied by a rejection based on 35 U.S.C. 112, first paragraph. The 35 U.S.C. 112, first paragraph, rejection should be set out as a separate rejection that incorporates by reference the factual basis and conclusions set forth in the 35 U.S.C. 101 rejection. The 35 U.S.C. 112, first paragraph, rejection should indicate that because the invention as claimed does not have utility, a person skilled in the art would not be able to use

> the invention as claimed, and as such, the claim is defective under 35 U.S.C. 112, first paragraph. A 35 U.S.C. 112, first paragraph, rejection should not be imposed or maintained unless an appropriate basis exists for imposing a rejection under 35 U.S.C. 101. **In other words, Office personnel should not impose a 35 U.S.C. 112, first paragraph, rejection grounded on a "lack of utility" basis unless a 35 U.S.C. 101 rejection is proper.** In particular, the factual showing needed to impose a rejection under 35 U.S.C. 101 must be provided if a rejection under 35 U.S.C. 112, first paragraph, is to be imposed on "lack of utility" grounds.
>
> ***
>
> To avoid confusion during examination, any rejection under 35 U.S.C. 112, first paragraph, based on grounds other than "lack of utility" should be imposed separately from any rejection imposed due to "lack of utility" under 35 U.S.C. 101 and 35 U.S.C. 112, first paragraph. *M.P.E.P.* § 2107.01 IV (emphasis added).

2. ***Enablement – Burden of Proof***

In order to make an enablement rejection, the PTO has the initial burden to establish a reasonable basis to question the enablement provided for the claimed invention. *See M.P.E.P.* § 2164.04. A specification teaching how to make and use the claimed subject matter must be taken as being in compliance with the enablement requirement unless there is a reason to doubt the objective truth of the statements contained therein which are relied on for enabling support. *Id.* It is incumbent for the PTO "to explain why it doubts the truth or accuracy of any statement in a supporting disclosure and to back up assertions of its own with acceptable evidence or reasoning which is inconsistent with the contested statement." *Id.* (quoting *In re Marzocchi*, 439 F.2d 220, 224, 169 U.S.P.Q. 367, 370 (C.C.P.A. 1971). This can be done "by making specific findings of fact, supported by the evidence, and then drawing conclusions based on these findings of fact." *Id.*

3. ***Enablement – Standard of Proof***

Once the examiner has weighed all the evidence and established a reasonable basis to question the enablement provided for the claimed invention, the burden falls on the applicant to present persuasive arguments, supported by suitable proofs where necessary, that one skilled in the art would be able to make and use the claimed invention using the application as a guide. *See*

M.P.E.P. § 2164.05. "The evidence provided by applicant **need not be conclusive but merely convincing** to one skilled in the art." *Id.* (bold emphasis added, underline in original). "A declaration or affidavit is, itself, evidence that must be considered." *Id.* (emphasis in original).

The examiner must then "weigh all the evidence before him or her, including the specification and any new evidence supplied by applicant with evidence and/or sound scientific reasoning previously presented in the rejection and decide whether the claimed invention is enabled." *Id.* "The examiner should **never** make the determination based on personal opinion. The determination should always be based on the weight of all the evidence." *Id.* (emphasis in original).

4. ***Appellants' Specification Teaches How to Make and Use the Claimed Subject Matter***

The specification enables one skilled in the art to make and use the full scope of the claims without undue experimentation. The claimed subject matter relates to polypeptides of SEQ ID NO:110 and the polypeptide encoded by ATCC deposit 203535, and polypeptides which are at least 95% identical to those polypeptides and which can be used to make antibodies that specifically detect PRO1753 in esophageal tissue.

The specification discloses how to make the claimed polypeptides, for example in paragraphs [0283]-[0315] and Examples 6-9 (¶¶ [0453]-[0492]). In addition, methods for making polypeptides which are at least 95% identical to SEQ ID NO:110 by making substitutions or deletions are also disclosed in the specification and were well known in the art. *See e.g., Specification* at paragraphs [0256]-[0271]. Methods for making and testing antibodies for specificity were well known in the art, and are disclosed in the specification, including paragraphs [0361]-[0379] and Example 10 (¶¶ [0493]-[0499]) of the specification, which specifically describes the preparation of antibodies that bind PRO polypeptides. In addition, the specification discloses that antibodies to claimed polypeptides can be used in diagnostic assays to detect the expression of PRO1753 in specific types of tissue. *See e.g., Specification* at [0407].

In light of the differential expression of the nucleic acid encoding the PRO1753 polypeptide in esophageal tumors compared to normal esophageal, one of skill in the art would

expect the PRO1753 polypeptide to be differentially expressed in these tumors as well. Therefore, given the teaching in the specification on how to make and use the claimed polypeptides to detect expression of PRO1753 in specific tissues, one of skill in the art would be enabled to practice the claimed invention without undue experimentation.

Because Appellants' specification teaches how to make and use the claimed subject matter, it must be taken as being in compliance with the enablement requirement unless there is a reason to doubt the objective truth of the statements contained therein which are relied on for enabling support. *See M.P.E.P.* § 2164.04.

5. *The Examiner's Arguments Fail to Establish a Reasonable Basis to Question the Enablement Provided for the Claimed Invention in the Specification*

The PTO has the initial burden to establish a reasonable basis to question the enablement provided for the claimed invention. *See M.P.E.P.* § 2164.04. It is incumbent for the PTO "to explain why it doubts the truth or accuracy of any statement in a supporting disclosure and to back up assertions of its own with acceptable evidence or reasoning which is inconsistent with the contested statement." *Id.* (quoting *In re Marzocchi*, 439 F.2d 220, 224, 169 U.S.P.Q. 367, 370 (C.C.P.A. 1971). This can be done "by making specific findings of fact, supported by the evidence, and then drawing conclusions based on these findings of fact." *Id.*

First, the Examiner has rejected pending Claims 6-8 and 11-17 as not being enabled because the claimed invention is not supported by either a specific or substantial asserted utility or a well-established utility. *See First Office Action* at 24. For the reasons discussed above, the claimed polypeptides possess a specific, substantial and credible utility, and to the extent that the enablement rejection is based on a lack of utility, it should be reversed.

Second, the Examiner has separately rejected pending Claims 12-17 under 35 U.S.C. 112, first paragraph, on the assertion that they contain subject matter which was not described in the specification in such a way as to enable one skilled in the art to make and/or use the invention. *Office Action* at 25.

The Examiner makes several arguments in rejecting claims which recite the limitation "wherein said isolated polypeptide is more highly expressed in esophageal tumor tissue

compared to normal esophageal tissue, or wherein said isolated polypeptide is encoded by a polynucleotide that is more highly expressed in esophageal tumor tissue compared to normal esophageal tissue." Specifically, the Examiner states that regarding the limitation "wherein said isolated polypeptide is more highly expressed in esophageal tumor tissue compared to normal esophageal tissue," no information is provided "regarding the level of expression, activity, or role in cancer of the PRO1753 polypeptide and validation of its association with lung [*sic*] tumors." *Final Office Action* at 25. The Examiner repeats his arguments from the utility rejection, stating that he has cited countervailing evidence that protein levels are not always consistent with mRNA levels. *Id.* at 25-26. As to the limitation "wherein said isolated polypeptide is encoded by a polynucleotide that is more highly expressed in esophageal tumor tissue compared to normal esophageal tissue," the Examiner summarily states that this limitation "does not limit the claimed polypeptide." *Id.* at 25.

While Appellants do not accept these arguments, Claims 4 and 5 which recited the limitations at issue have been canceled, and Claim 12, which depended from Claim 4, has been amended to depend from Claim 6. Claims 12-13 as dependent from Claim 6 do not recite percent amino acid sequence identity as a limitation, nor do they recite any limitation regarding overexpression in esophageal tumors. These claims are directed to fusion peptides of the disclosed sequence, with or without the disclosed signal peptide. Therefore, the amendments to the pending claims have rendered these arguments moot.

With respect to Claims 14-17, which recite the limitation "wherein said isolated polypeptide or a fragment thereof can be used to generate an antibody which can be used to specifically detect the polypeptide of SEQ ID NO: 110 in esophageal tissue samples," the Examiner states that "[t]hese claims encompass any and all antigenically cross-reactive polypeptides possessing the recited percent identity, regardless of their biological activity." *Id.* at 26. The Examiner continues, arguing:

> If mere antigenic cross-reactivity were the test for enablement under § 112, Applicants could obtain patent rights that may confer power to block off whole areas of scientific development related to the biological activity of the polypeptide, for which Applicants have not provided any disclosure. It is entirely unclear why the disclosure of a single polypeptide, i.e., PRO1753, which is ideally suited to the making of antibodies to itself, would enable any and all antigenically

> cross-reactive polypeptides possessing the recited percent identity and possessing unknown and undisclosed biological activities, when the specification provides no disclosure of any biological activity. Therefore, the scope of enablement provided to the skilled artisan by the disclosure is not commensurate with the scope of protection sought by the claims. *Id.* at 26-27.

The standard for determining whether the specification meets the enablement requirement is to be evaluated based on whether or not the experimentation needed for one skilled in the art to practice the invention would be undue. *Mineral Separation v. Hyde,* 242 U.S. 261, 270 (1916); *In re Wands*, 858 F.2d 731, 737, 8 U.S.P.Q.2d 1400, 1404 (Fed. Cir. 1988); *M.P.E.P.* § 2164.01. Applicants submit that in view of the requirements of enablement under 35 U.S.C. §112, first paragraph, the Examiner has failed to establish a *prima facie* basis for rejecting Claims 14-17 as lacking enablement. The Examiner's unsupported conclusory statements fail to establish a reasonable basis to question the enablement provided for the claimed invention. *See M.P.E.P.* § 2164.04.

It is incumbent for the PTO "to explain why it doubts the truth or accuracy of any statement in a supporting disclosure and to back up assertions of its own with acceptable evidence or reasoning which is inconsistent with the contested statement." *Id.* (quoting *In re Marzocchi*, 439 F.2d 220, 224, 169 U.S.P.Q. 367, 370 (C.C.P.A. 1971). This can be done "by making specific findings of fact, supported by the evidence, and then drawing conclusions based on these findings of fact." *Id.* The Examiner has failed to make any specific findings of fact, or back up his assertions with any acceptable evidence or reasoning.

In the present case, the Examiner reasons that "[i]f mere antigenic cross-reactivity were the test for enablement under § 112, Applicants could obtain patent rights that may confer power to block off whole areas of scientific development related to the biological activity of the polypeptide, for which Applicants have not provided any disclosure." This is not the test for enablement.

The M.P.E.P. states that "if any use is enabled when multiple uses are disclosed, the application is enabling for the claimed invention." *M.P.E.P.* § 2164.01(c) (emphasis added). As described above, the specification adequately discloses how to make and use the polypeptides of Claims 14-17. The Examiner has not alleged that undue experimentation would be required to

practice the claimed invention. Accordingly, it remains unquestioned that the claimed polypeptides have an enabled use.

Asserting that "[i]t is entirely unclear why the disclosure of a single polypeptide, ... would enable any and all antigenically cross-reactive polypeptides possessing the recited percent identity and possessing unknown and undisclosed biological activities, when the specification provides no disclosure of any biological activity," is also not a reason to reject the claimed polypeptides as lacking enablement. The subject matter of Claims 14-17 relates to isolated polypeptides with at least 95% identity to the disclosed polypeptides wherein the claimed polypeptides can be used to generate an antibody which can be used to specifically detect the polypeptide of SEQ ID NO: 110 in esophageal tissue samples. The Examiner has not offered any explanation of how the failure to disclose "any biological activity" results in one of skill in the art having to resort to undue experimentation to practice the claimed invention. Disclosure of a "biological activity" is not required for one of skill in the art to either make or use the claimed polypeptides.

Further, "[t]he presence of only one example should never be the sole reason for rejecting claims as being broader than the enabling disclosure, even though it is a factor to be considered along with all the other factors. To make a valid rejection, one must evaluate all the facts and evidence and state why one would not expect to be able to extrapolate that one example across the entire scope of the claims." *M.P.E.P.* § 2164.02. The Examiner provides no other basis for rejection of the Claims 14-17 aside from pointing to "disclosure of a single polypeptide." Accordingly, pointing to the disclosure of a single polypeptide, absent any other evidence, cannot support *prima facia* rejection of lack of enablement of the claimed polypeptides.

Notwithstanding the failure of the Examiner to provide sufficient evidence to support a *prima facie* rejection of Claims 14-17, as explained above, the specification teaches in detail how to make the claimed polypeptides, including variants thereof, and antibodies which specifically bind PRO1753. Likewise, as detailed above, the specification provides sufficient guidance as to how to use the claimed polypeptides. Thus, there is significant guidance how to make and use the claimed polypeptides. In addition, as the disclosure and references cited in the specification make clear, the production of polypeptides, polypeptide variants, and specific antibodies is a

predictable and well established aspect of the biological sciences. *See, e.g., In re Wands*, 858 F.2d 731, 8 U.S.P.Q. 2d 1400 (Fed. Cir. 1988) (reversing the Board's decision of non-enablement and holding that as of 1980, undue experimentation was not required to make high-affinity monoclonal antibodies to a target peptide).

In conclusion, the Examiner has failed to meet his burden to establish a reasonable basis to question the enablement provided for the claimed invention – conclusory statements are simply not sufficient.

6. *Appellants' Response and Grouping of Rejected Claim – Rejected Claims 6-8 and 11-17 are Enabled*

Appellants submit that for the reasons stated above, the rejected claims are enabled by the specification as filed. The Examiner has failed to "to explain why [he] doubts the truth or accuracy of any statement in a supporting disclosure and to back up assertions of [his] own with acceptable evidence or reasoning which is inconsistent with the contested statement." *M.P.E.P.* § 2164.04. Instead, the Examiner impermissibly relies on his own personal opinion that "it is unclear" how the disclosure enables the claimed subject matter. *Id.* ("The examiner should **never** make the determination [of whether a claim is enabled] based on personal opinion. The determination should always be based on the weight of all the evidence.") (some emphasis supplied).

For purposes of the enablement rejection, Claims 6-8 and 11-13 can be considered as a group, Claims 14 and 16-17 can be considered as a group, and Claim 15 should be considered individually.

a. Claims 6-8 and 11-13 are enabled

Claims 6-8 and 11-13 are enabled for the reasons discussed above. The scope of these claims is narrow, and because the sequences of SEQ ID NO:110, the signal peptide, and ATCC deposit 203535 are explicitly disclosed in the specification, no experimentation of any kind is required to make the claimed polypeptides. One of skill in the art would clearly be able to use these polypeptides to make antibodies which are specific to these polypeptides, such that the

level of expression of these polypeptides could be assessed in esophageal tissue. The only question is whether the use of these polypeptides to make antibodies to detect their expression level is a substantial and specific utility. For the reasons discussed at length above, Appellants believe that differential expression of the PRO1753 mRNA in esophageal tumors provides the claimed polypeptides with a substantial and specific utility. Therefore, Claims 6-8 and 11-13 are enabled.

b. Claims 14 and 16-17 are enabled

Claims 14 and 16-17 are enabled for the reasons discussed above. Because SEQ ID NO:110 is 545 amino acids long, a polypeptide which is at least 95% identical can only have approximately 27 deletions or substitutions. While some experimentation will be required to make these polypeptides that are not identical to SEQ ID NO:110, any such experimentation is routine in the art and will not be undue. One of skill in the art would clearly be able to use these polypeptides to make antibodies which are specific to SEQ ID NO:110, such that the expression level of PRO1753 can be assessed in esophageal tissue. The only question is whether the use of these polypeptides to make antibodies to detect the expression level of PRO1753 is a substantial and specific utility. For the reasons discussed at length above, Appellants believe that it is, and therefore, Claims 14 and 16-17 are enabled.

c. Claim 15 is enabled

Claim 15 is enabled for the reasons discussed above. The scope of this claim is narrower than that of Claim 14 since a polypeptide which is at least 99% identical can only have approximately 5 deletions or substitutions. As a result, less experimentation will be required to make these polypeptides, although any experimentation remains routine. One of skill in the art would clearly be able to use these polypeptides to make antibodies which are specific to SEQ ID NO:110, and for the reasons discussed above, Appellants believe that such use is substantial and specific. Therefore, Claim 15 is enabled.

7. ***Enablement – Conclusion***

For the reasons discussed above, the specification enables the scope of the claimed invention such that one skilled in the art could make and use the claimed invention without undue experimentation. The Examiner has offered only conclusory statements, and has failed to back up his assertions "with acceptable evidence or reasoning which is inconsistent with the contested statement [of enablement]." *Id.* (quoting *In re Marzocchi*, 439 F.2d 220, 224, 169 U.S.P.Q. 367, 370 (C.C.P.A. 1971). Therefore, the Examiner has failed to meet his initial burden to establish a reasonable basis to question the enablement provided for the claimed invention. *See M.P.E.P.* § 2164.04.

And even if the Examiner has met his burden, Appellants have presented persuasive arguments, supported by the evidence discussed above with respect to utility, that one skilled in the art would be able to make and use the claimed invention using the application as a guide. Appellants remind the Board that "[t]he evidence provided by applicant need not be conclusive but merely convincing to one skilled in the art." *M.P.E.P.* § 2164.05 (emphasis in original).

Appellants submit that rejected Claims 6-8 and 11-13 should be considered as a group, Claims 14 and 16-17 should be considered as a group, and Claim 15 should be considered individually. This is because the scope of the subject matter in each of the groups differs, and therefore varying amounts of experimentation will be required to make and use each of the groups. However, Appellants submit that any experimentation would be routine for those of skill in the art.

Considering all of the evidence provided by the Appellants to establish their asserted utility, along with the disclosure in the specification, the Board should find that Appellants have established that one of skill in the art would be able to make and use the claimed invention without undue experimentation, and the Examiner's rejection of the pending claims as lacking an enabling disclosure should be reversed.

D. Written Description Rejection – Detailed Argument

The third issue before the Board is whether the claimed subject matter is described in the specification in such a way as to reasonably convey to one skilled in the art that the inventors had

possession of the claimed invention at the time the application was filed. The Examiner has rejected pending Claims 12-17 under 35 U.S.C. §112, first paragraph, as lacking an adequate written description. *Office Action* at 27.

The Examiner has failed to meet his initial burden of rebutting the presumption that the written description is adequate because the Examiner's only arguments directed to the Claims 12-17 are the same unsupported conclusory arguments made to support the enablement rejection of the pending claims. For the reasons detailed below, Appellants submit that Claims 12-17 are adequately described such that one of skill in the art would recognize that the inventors had possession of the claimed invention at the time the application was filed.

1. *Written Description – Legal Standard*

The well-established test for sufficiency of support under the written description requirement of 35 U.S.C. §112, first paragraph is stated by the Court in *Vas-Cath, Inc. v. Mahurkar*, 935 F.2d 1555, 19 U.S.P.Q.2d 1111 (Fed. Cir. 1991):

> "Although [the applicant] does not have to describe exactly the subject matter claimed, . . . the description must clearly allow persons of ordinary skill in the art to recognize that [he or she] invented what is claimed." "The test for sufficiency of support in a parent application is whether the disclosure of the application relied upon 'reasonably conveys to the artisan that the inventor had possession at that time of the later claimed subject matter.'" *Vas-Cath, Inc. v. Mahurkar*, 935 F.2d at 1562-63, 19 U.S.P.Q.2d at 1116 (citations omitted)

2. *Written Description – Burden of Proof*

The M.P.E.P. states that "[a] description as filed is presumed to be adequate, unless or until sufficient evidence or reasoning to the contrary has been presented by the examiner to rebut the presumption. See, e.g., *In re Marzocchi*, 439 F.2d 220, 224, 169 U.S.P.Q. 367, 370 (C.C.P.A. 1971)." *M.P.E.P.* § 2163.04 (emphasis added). Therefore "[t]he examiner has the initial burden of presenting by a preponderance of evidence why a person skilled in the art would not recognize in an applicant's disclosure a description of the invention defined by the claims. *Wertheim*, 541 F.2d at 263, 191 U.S.P.Q. at 97." *Id.* Only then does the Applicant need to respond to the Examiner's arguments.

3. *Written Description – Standard of Proof*

The adequacy of written description support is a factual issue and is to be determined on a case-by-case basis. *See e.g.*, *Vas-Cath, Inc. v. Mahurkar*, 935 F.2d at 1563, 19 U.S.P.Q.2d at 1116 (Fed. Cir. 1991) (emphasis added). The factual determination in a written description analysis depends on the nature of the invention and the amount of knowledge imparted to those skilled in the art by the disclosure. *Union Oil v. Atlantic Richfield Co.*, 208 F.3d 989, 996 (Fed. Cir. 2000). As with the utility requirement, the evidentiary standard to be used throughout *ex parte* examination in setting forth a rejection is a preponderance of the evidence, or "more likely than not" standard. *In re Oetiker*, 977 F.2d 1443, 1445, 24 U.S.P.Q.2d 1443, 1444 (Fed. Cir. 1992).

4. *The Examiner's Arguments Fail to Overcome the Presumption that the Claimed Invention is Adequately Described in the Specification*

To overcome the presumption that the claimed subject matter is adequately described, the Examiner must present "by a preponderance of evidence why a person skilled in the art would not recognize in an applicant's disclosure a description of the invention defined by the claims." *M.P.E.P.* § 2163.04 (emphasis added). To support his rejection of pending Claims 12-17, the Examiner has merely repeated, nearly verbatim, the same unsupported conclusory arguments made in support of his enablement rejection.

In response to claims reciting the limitation "wherein said isolated polypeptide is more highly expressed in esophageal tumor tissue compared to normal esophageal tissue," the Examiner repeats his assertion that no information is provided "regarding the level of expression, activity, or role in cancer of the PRO1753 polypeptide." *Final Office Action* at 27. The Examiner also repeats his arguments from the utility rejection, stating that he has cited countervailing evidence that protein levels are not always consistent with mRNA levels. *Id.* As to the limitation "wherein said isolated polypeptide is encoded by a polynucleotide that is more highly expressed in esophageal tumor tissue compared to normal esophageal tissue," the

Examiner summarily states that this limitation "does describe the claimed polypeptides." *Id.* at 27.

While Appellants do not agree with these assertions, Claims 4 and 5 which recited the limitations at issue have been canceled, and Claim 12, which depended from Claim 4, has been amended to depend from Claim 6. Claims 12-13 as dependent from Claim 6 do not recite percent amino acid sequence identity as a limitation, nor do they recite any limitation regarding overexpression in esophageal tumors. These claims are directed to fusion peptides of the disclosed sequence, with or without the disclosed signal peptide. Therefore, the amendments to the pending claims have rendered these arguments moot. In the absence of any other arguments as to "why one of skill in the art would not recognize in an applicant's disclosure a description of the invention defined by the claims," the Examiner has failed to rebut the presumption that the specification satisfies the written description requirement for Claims 12-13. *See M.P.E.P.* § 2163.04.

With respect to Claims 14-17, which recite the limitation "wherein said isolated polypeptide or a fragment thereof can be used to generate an antibody which can be used to specifically detect the polypeptide of SEQ ID NO: 110 in esophageal tissue samples," the Examiner repeats his assertion that "[t]hese claims encompass any and all antigenically cross-reactive polypeptides possessing the recited percent identity, regardless of their biological activity." *Id.* at 28. The Examiner continues, arguing:

> Applicants have not described the biologic activity of the PRO1753 polypeptide or any of its variants. It is entirely unclear why the disclosure of a single polypeptide, i.e., PRO1753, which is ideally suited to the making of antibodies to itself, would describe any and all antigenically cross-reactive polypeptides possessing the recited percent identity and possessing unknown and undisclosed biological activities, when the specification does not describe any biological activity. *Id.* at 28.

The Examiner concludes that as a result, the claimed subject matter is not adequately described. *Id.*

As noted above, "[a] description as filed is presumed to be adequate, unless or until sufficient evidence or reasoning to the contrary has been presented by the examiner to rebut the presumption." *M.P.E.P.* § 2163.04 (emphasis added). Therefore "[t]he examiner has the initial

burden of presenting by a preponderance of evidence why a person skilled in the art would not recognize in an applicant's disclosure a description of the invention defined by the claims." *Id.*

The Examiner has not provided any reasoning or evidence as to how the absence of the disclosure of "biological activity" results in an inadequate description of the subject matter of Claims 14-17. The claimed subject matter relates to polypeptides that have at least 95% sequence identity to SEQ ID NO:110, and can be used to generate an antibody which can be used to specifically detect the polypeptide of SEQ ID NO: 110 in esophageal tissue samples. Appellants fail to see how any "biological activity" of the claimed polypeptides, aside from being able to generate an antibody which can specifically detect the polypeptide of SEQ ID NO: 110, is at all relevant to an adequate description of the claimed polypeptides which are not claimed on the basis of any "biological activity." In the absence of any other arguments as to why one of skill in the art would not recognize a description of the claimed invention in Appellants' disclosure, the Examiner has failed to rebut the presumption that the specification satisfies the written description requirement for Claims 14-17. *See M.P.E.P.* § 2163.04.

5. *Appellants' Response and Grouping of Claims – Rejected Claims 12-17 are Adequately Described*

The adequacy of written description support is a factual issue and is to be determined on a case-by-case basis. *See e.g., Vas-Cath, Inc. v. Mahurkar*, 935 F.2d at 1563, 19 U.S.P.Q.2d at 1116 (Fed. Cir. 1991) (emphasis added). The factual determination in a written description analysis depends on the nature of the invention and the amount of knowledge imparted to those skilled in the art by the disclosure. *Union Oil v. Atlantic Richfield Co.*, 208 F.3d 989, 996 (Fed. Cir. 2000).

As discussed above, the Examiner must address each of the pending claims on a case-by-case basis. This is because the genuses encompassed by the claims differ, and therefore whether or not the specification supports the claimed genus depends on the claim at issue. Appellants hereby request that the Board consider the following groupings of Claims 12-17 with respect to the written description requirement.

a. Rejected Claims 12-13 are adequately described

Claims 12-13, which depend from Claim 6, are adequately described by the specification. Claim 6 is directed to an isolated polypeptide comprising the amino acid sequence of the polypeptide of SEQ ID NO:110, the amino acid sequence of the polypeptide of SEQ ID NO:110 lacking its associated signal peptide, or the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203535.

As stated above, the Examiner provides no basis for rejecting either of Claims 12-13 because the Examiner's arguments are directed at claims reciting the limitation "wherein said isolated polypeptide is encoded by a polynucleotide that is more highly expressed in esophageal tumor tissue compared to normal esophageal tissue," or "wherein said isolated polypeptide or a fragment thereof can be used to generate an antibody which can be used to specifically detect the polypeptide of SEQ ID NO: 110 in esophageal tissue samples." *See Final Office Action* at 27-29. As amended, Claims 12-13 depend from Claim 6, which does not recite the objected to limitation.

Regardless of any lack of reasoning provided by the Examiner, Appellants assert that each recited element of Claim 6 is explicitly disclosed in the specification, either in writing (*see, e.g.*, *Specification* at Figure 110) or by virtue of a biological deposit. Accordingly, there can be no basis for holding that Claim 6 is not adequately described. Likewise, Claims 12-13, which are drawn to chimeric polypeptides comprising the polypeptide of Claim 6, are also fully described by the specification.

The Examiner does not contest the written description support for any embodiment recited in Claims 12-13. Therefore the Examiner has failed to meet his "initial burden of presenting by a preponderance of evidence why a person skilled in the art would not recognize in an applicant's disclosure a description of the invention defined by the claims." *M.P.E.P.* § 2163.04. As such, the Board should reverse the Examiner's rejection of Claims 12-13 under 35 U.S.C. § 112, first paragraph, for lack of written description.

b. Rejected Claim 14, 16 and 17 are adequately described

Claims 14, 16 and 17 are adequately described by the specification. Claim 14 is directed to an isolated polypeptide having at least 95% amino acid sequence identity to the amino acid sequence of the polypeptide SEQ ID NO:110, the amino acid sequence of the polypeptide of SEQ ID NO:110 lacking its associated signal peptide, or the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203535; wherein said isolated polypeptide or a fragment thereof can be used to generate an antibody which can be used to specifically detect the polypeptide of SEQ ID NO:110 in esophageal tissue samples. Claims 16 and 17 ultimately depend from Claim 14.

Appellants maintain that there is no substantial variation within the species which fall within the scope of the rejected claims, which require at least 95% amino acid sequence identity to SEQ ID NO:110 and can be used to generate antibodies which specifically detect the polypeptide of SEQ ID NO:110 in esophageal tissue samples. As such, Appellants were in possession of the common attributes or features of the claimed subject matter.

The rejected claims are analogous to the claims discussed in Example 14 of the written description training materials available on the PTO's website. In Example 14, the written description requirement was found to be satisfied for claims directed to polypeptides with 95% homology to a disclosed sequence that also possess a recited catalytic activity, where procedures for making variant proteins were routine in the art and the specification provided an assay for detecting the recited catalytic activity of the protein. This disclosure satisfies the written description requirement even though the applicant had disclosed only a single species and had not made any variants. The Guidelines state that "[t]he single species disclosed is representative of the genus because all members have at least 95% structural identity with the reference compound and because of the presence of an assay which applicant provided for identifying all of the at least 95% identical variants of SEQ ID NO: 3 which are capable of the specified catalytic activity."

Similarly, the pending claims also have at least 95% sequence identity to the disclosed sequence, and must be able to generate antibodies which specifically detect the polypeptide of SEQ ID NO:110 in esophageal tissue samples. As in Example 14, at the time of the effective

filing date of the instant application, it was well known in the art how to make polypeptides with at least 95% amino acid sequence identity to the disclosed sequences. *See, e.g., Specification* at ¶¶ [0256]-[0271]. In addition, the specification discloses in detail how to make antibodies which specifically detect a particular PRO polypeptide, and how to use them to detect the PRO polypeptide in a particular tissue. *See, e.g., Specification* ¶¶ [0363]-[0379], [0407], and [0493]-[0499]. Like a particular catalytic activity, the function of being useful to produce an antibody specific to SEQ ID NO:110 is directly related to the structure of the claimed polypeptides. Thus, like Example 14, the genus of polypeptides that have at least 95% amino acid sequence identity to the disclosed sequences and possess the described functional activity are adequately described.

Claims 16 and 17, drawn to particular embodiments of Claim 14, are also fully described by the specification. The Examiner does not contest the written description support for any embodiment recited in Claims 16-17.

The Examiner asserts that the present claims are not analogous to those discussed in Example 14 of the written description guidelines because the specification does not describe any biological activity of the claimed polypeptides and because the claims are not limited to any specific "biological activity" of the claimed polypeptides. *Final Office Action* at 28-29.

Appellants submit that the applicability of Example 14 is not limited to polypeptides for which the biological function is known and recited, but extends to all situations where the polypeptide is useful and there is no substantial variation within the species encompassed by the claims. The purpose of the recited catalytic activity in the example is to limit the amount of structural variation within the species. The commentary in the Guidelines states that the description of an assay to detect variants which have the recited activity, along with 95% homology, is sufficient to satisfy the written description requirement.

Similarly, in the instant case, Claims 14-17 must share a particular "biological activity" which restricts the amount of permissible structural variation within the species – the claimed polypeptides must be useable to generate an antibody which can be used to specifically detect the polypeptide of SEQ ID NO: 110 in esophageal tissue samples. This limitation combined with the disclosure of how to make and test the recited antibodies generated from the claimed polypeptides, along with the requirement of least 95% amino acid sequence identity, results in

claimed subject matter where there is no substantial variation within the species encompassed by the claims. Accordingly, Appellants maintain that the pending claims are analogous to the claims in Example 14.

As for the Examiner's conclusory, and unsupported statement that "[i]t is entirely unclear why the disclosure of a single polypeptide, i.e., PRO1753, ...would describe any and all antigenically cross-reactive polypeptides possessing the recited percent identity," the basic premise that a large genus can not be adequately described by a single species is simply wrong. In a recent Federal Circuit decision, *In re Wallach*, 378 F.3d 1330, 1333-34 (Fed. Cir. 2004), the Court stated:

> [W]e agree with Appellants that the state of the art has developed such that the complete amino acid sequence of a protein may put one in possession of the genus of DNA sequences encoding it, and that one of ordinary skill in the art at the time the '129 application was filed may have therefore been in possession of the entire genus of DNA sequences that can encode the disclosed partial protein sequence, even if individual species within that genus might not have been described or rendered obvious. ... A claim to the genus of DNA molecules complementary to the RNA having the sequences encompassed by that formula, even if defined only in terms of the protein sequence that the DNA molecules encode, while containing a large number of species, is definite in scope and provides the public notice required of patent applicants.
>
> ***
>
> Moreover, we see no reason to require a patent applicant to list every possible permutation of the nucleic acid sequences that can encode a particular protein for which the amino acid sequence is disclosed, given the fact that it is, as explained above, a routine matter to convert back and forth between an amino acid sequence and the sequences of the nucleic acid molecules that can encode it. *Id.* (emphasis added).

The Court did not require the applicants in *Wallach* to actually make or individually describe all of the *vast* number of sequences which encode the disclosed protein sequence. This is in spite of the fact that only a single protein sequence was disclosed, and the encompassed genus was enormous due to codon degeneracy in the genetic code – even the most skilled artisan could not individually envision the detailed chemical structure of the nucleic acids encompassed by the claimed genus. The Court reasoned that because it is routine to convert between amino

acid sequences and nucleic acid sequences, disclosure of a single amino acid sequence was sufficient to place the applicants in possession of the enormous genus of nucleic acids which could encode the sequence.

The facts in *Wallach* are very similar to the instant case. Here, Appellants have disclosed SEQ ID NO:110, and claim polypeptides which are at least 95% identical to it and have the functional limitation of the ability to generate antibodies which can be used to specifically detect SEQ ID NO:110 in esophageal tissue samples. As discussed above, it is routine in the art to create polypeptides which have at least 95% sequence identity to SEQ ID NO:110 – it is just as predictable and easy as creating all of the nucleic acids which encode a particular amino acid sequence. Similarly, it is well within the knowledge of those skilled in the art how to determine which polypeptides can be used to make the recited antibodies. The predictability of this structure/function combination is sufficient to place the claimed subject matter in the possession of the Appellants, and thus the claimed polypeptides are adequately described. The *Wallach* opinion makes clear that there is no need to literally describe more than a single species to adequately describe a large genus where one of skill in the art recognizes that the disclosed species puts the applicant in possession of the claimed genus.

In conclusion, the Examiner has failed to meet his "initial burden of presenting by a preponderance of evidence why a person skilled in the art would not recognize in an applicant's disclosure a description of the invention defined by the claims." *M.P.E.P.* § 2163.04. And even if he has met this burden, Appellants responded with arguments and evidence which are sufficient to overcome the Examiner's conclusory statements and unsupported opinions. As such, the Board should reverse the Examiner's rejection of Claims 14 and 16-17 under 35 U.S.C. § 112, first paragraph, for lack of written description.

c. Rejected Claim 15 is adequately described

For the reasons discussed above regarding Claims 14 and 16-17, Appellants believe that Claim 15 is also adequately described. However, because SEQ ID NO:110 is 545 amino acids long, a polypeptide which is at least 99% identical can only have approximately 5 deletions or substitutions. As a result, the genus of polypeptides encompassed by Claim 15 is smaller than

that of Claim 14, and the Board should consider the adequacy of the written description for this claim independently of the other claims. Given the lack of any substantial arguments by the Examiner to support his rejection, as well as his failure to address Claim 15 directly, the Board should reverse the Examiner's rejection of Claim 15 under 35 U.S.C. § 112, first paragraph, for lack of written description.

6. *Written Description – Conclusion*

In conclusion, the Board should reverse the Examiner's written description rejection of Claims 12-17 because the Examiner has failed to rebut the presumption that the claims are adequately described, as he has failed to provide any evidence or reasoning to support the rejection: "[a] description as filed is presumed to be adequate, unless or until sufficient evidence or reasoning to the contrary has been presented by the examiner to rebut the presumption." *M.P.E.P.* § 2163.04 (emphasis added).

And even if the Examiner's arguments are sufficient to rebut the presumption, Appellants submit that they have satisfied the written description requirement for rejected Claims 12-17 based on the actual reduction to practice of SEQ ID NO:110, by specifying a high level of amino acid sequence identity, and by describing how to make and use antibodies to the disclosed sequence. These facts are directly analogous to those of Example 14 of the Written Description Guidelines published by the PTO. In addition, like *In re Wallach*, the description of the single species SEQ ID NO:110 is sufficient to place the Appellants' in possession of the claimed genus because those of skill in the art recognize the correlation between polypeptide structure and the ability to generate specific antibodies. Appellants submit that the instant disclosure allows one of skill in the art to "recognize that the applicant was in possession of the necessary common attributes or features of the elements possessed by the members of the genus." Hence, Appellants respectfully request that the Board reverse the Examiner's written description rejection of Claims 12-17 under 35 U.S.C. §112, first paragraph.

E. Conclusion

In view of the arguments presented above, Appellants submit that the specification as filed provides a specific, substantial and credible utility for the claimed polypeptides, and that the claimed subject matter is enabled and adequately described by the specification. Appellants therefore respectfully request that the Board reverse the rejections of the pending claims as lacking utility under 35 U.S.C. §101, and as not being enabled or adequately described under 35 U.S.C. §112.

Please charge any additional fees, including any fees for additional extension of time, or credit overpayment to Deposit Account No. 11-1410.

Respectfully submitted,

KNOBBE, MARTENS, OLSON & BEAR, LLP

Dated: Dec. 20, 2005

By: _____

AnneMarie Kaiser
Registration No. 37,649
Attorney of Record
Customer No. 30,313
(619) 235-8550

2153696
120105

VIII. APPENDIX A – Claims on Appeal

1-5. (Canceled).

6. An isolated polypeptide comprising:

(a) the amino acid sequence of the polypeptide of SEQ ID NO: 110;

(b) the amino acid sequence of the polypeptide of SEQ ID NO: 110, lacking its associated signal peptide; or

(c) the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203535.

7. The isolated polypeptide of Claim 6 comprising the amino acid sequence of the polypeptide of SEQ ID NO: 110.

8. The isolated polypeptide of Claim 6 comprising the amino acid sequence of the polypeptide of SEQ ID NO: 110, lacking its associated signal peptide.

9-10. (Canceled).

11. The isolated polypeptide of Claim 6 comprising the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203535.

12. A chimeric polypeptide comprising a polypeptide according to Claim 6 fused to a heterologous polypeptide.

13. The chimeric polypeptide of Claim 12, wherein said heterologous polypeptide is a tag polypeptide or an Fc region of an immunoglobulin.

14. An isolated polypeptide having at least 95% amino acid sequence identity to:

(a) the amino acid sequence of the polypeptide of SEQ ID NO: 110;

(b) the amino acid sequence of the polypeptide of SEQ ID NO: 110, lacking its associated signal peptide; or

(c) the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203535;

wherein said isolated polypeptide or a fragment thereof can be used to generate an antibody which can be used to specifically detect the polypeptide of SEQ ID NO: 110 in esophageal tissue samples.

15. The isolated polypeptide of Claim 14 having at least 99% amino acid sequence identity to:

(a) the amino acid sequence of the polypeptide of SEQ ID NO: 110;

(b) the amino acid sequence of the polypeptide of SEQ ID NO: 110, lacking its associated signal peptide; or

(c) the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203535;

wherein said isolated polypeptide or a fragment thereof can be used to generate an antibody which can be used to specifically detect the polypeptide of SEQ ID NO: 110 in esophageal tissue samples.

16. A chimeric polypeptide comprising a polypeptide according to Claim 14 fused to a heterologous polypeptide.

17. The chimeric polypeptide of Claim 16, wherein said heterologous polypeptide is a tag polypeptide or an Fc region of an immunoglobulin.

IX. APPENDIX B – Evidence

Attached hereto is a copy of the evidence cited in Appellants' Brief. The list of evidence below is accompanied by a statement setting forth where in the record that evidence was entered into the record by the Examiner.

Tab	Reference	Submitted	Entered
1	Hu *et al.* (J. Proteome Res. (2003) 2(4):405-12)		Cited by Examiner in the final Office Action
2	Wang (Trends Pharmacol. Sci., (1996) 17(8):276-9)		Cited by Examiner in the final Office Action
3	Haynes *et al.* (Electrophoresis, (1998) 19(11):1862-71)		Cited by Examiner in the final Office Action
4	Gygi *et al.* (Mol. and Cell. Bio., (1999).		Cited by Examiner in the final Office Action
5	Hancock (J. Proteome Res., (2004) 3(4):685)		Cited by Examiner in the final Office Action
6	Declaration of J. Christopher Grimaldi	Originally Submitted on April 29, 2005 with Appellants' Amendment and Response to Office Action as Exhibit 1	Entered by Examiner in final Office Action
7	Second Declaration by J. Christopher Grimaldi	Originally Submitted on April 29, 2005 with Appellants' Amendment and Response to Office Action as Exhibit 5	Entered by Examiner in final Office Action
8	Declaration of Paul Polakis, Ph.D.	Originally Submitted on April 29, 2005 with Appellants' Amendment and Response to Office Action as Exhibit 6	Entered by Examiner in final Office Action
9	Bruce Alberts, *et al.*, Molecular Biology of the Cell (3rd ed. 1994)	Originally Submitted on April 29, 2005 with Appellants' Amendment and Response to Office Action as Exhibit 7	Entered by Examiner in final Office Action
10	Bruce Alberts, *et al.*, Molecular Biology of the Cell (4th ed. 2002)	Originally Submitted on April 29, 2005 with Appellants' Amendment and Response to Office Action as Exhibit 8	Entered by Examiner in final Office Action

11	Genes VI, (Benjamin Lewin, Genes VI (1997)	Originally Submitted on April 29, 2005 with Appellants' Amendment and Response to Office Action as Exhibit 9	Entered by Examiner in final Office Action
12	Zhigang *et al.*, World Journal of Surgical Oncology 2:13, 2004	Originally Submitted on April 29, 2005 with Appellants' Amendment and Response to Office Action as Exhibit 10	Entered by Examiner in final Office Action
13	Meric *et al.*, Molecular Cancer Therapeutics, vol. 1, 971-979 (2002)	Originally Submitted on April 29, 2005 with Appellants' Amendment and Response to Office Action as Exhibit 11	Entered by Examiner in final Office Action

X. APPENDIX C – Related Proceedings

None – There are no decisions rendered by a court or the Board in any related proceedings identified above.

research articles Journal of proteome research

Analysis of Genomic and Proteomic Data Using Advanced Literature Mining

Yanhui Hu, Lisa M. Hines, Haifeng Weng, Dongmei Zuo, Miguel Rivera, Andrea Richardson, and Joshua LaBaer*

Institute of Proteomics, Harvard Medical School–BCMP, 240 Longwood Avenue, Boston, Massachusetts 02115

Received March 13, 2003

High-throughput technologies, such as proteomic screening and DNA micro-arrays, produce vast amounts of data requiring comprehensive analytical methods to decipher the biologically relevant results. One approach would be to manually search the biomedical literature; however, this would be an arduous task. We developed an automated literature-mining tool, termed MedGene, which comprehensively summarizes and estimates the relative strengths of all human gene–disease relationships in Medline. Using MedGene, we analyzed a novel micro-array expression dataset comparing breast cancer and normal breast tissue in the context of existing knowledge. We found no correlation between the strength of the literature association and the magnitude of the difference in expression level when considering changes as high as 5-fold; however, a significant correlation was observed ($r = 0.41$; $p = 0.05$) among genes showing an expression difference of 10-fold or more. Interestingly, this only held true for estrogen receptor (ER) positive tumors, not ER negative. MedGene identified a set of relatively understudied, yet highly expressed genes in ER negative tumors worthy of further examination.

Keywords: bioinformatics • micro-array • text mining • gene-disease association • breast cancer

Introduction

At its current pace, the accumulation of biomedical literature outpaces the ability of most researchers and clinicians to stay abreast of their own immediate fields, let alone cover a broader range of topics. For example, to follow a single disease, e.g., breast cancer, a researcher would have had to scan 130 different journals and read 27 papers per day in 1999.[1] This problem is accentuated with high-throughput technologies such as DNA micro-arrays and proteomics, which require the analysis of large datasets involving thousands of genes, many of which are unfamiliar to a particular researcher. In any microarray experiment, thousands of genes may demonstrate statistically significant expression changes, but only a fraction of these may be relevant to the study. The ability to interpret these datasets would be enhanced if they could be compared to a comprehensive summary of what is known about all genes. Thus, there is a need to summarize existing knowledge in a format that allows for the rapid analysis of associations between genes and diseases or other specific biological concepts.

One solution to this problem is to compile structured digital resources, such as the Breast Cancer Gene Database[1] and the Tumor Gene Database.[2] However, as these resources are hand-curated, the labor-intensive review process becomes a rate-limiting step in the growth of the database. As a result, these databases have a limited scale and the genes are not selected in a systematic fashion.

An alternative approach is automated text mining; a method which involves automated information extraction by searching documents for text strings and analyzing their frequency and context. This approach has been used successfully in several instances for biological applications. In most cases, it has been applied to extract information about the relationships or interactions that proteins or genes have with one another, in the literature or by functional annotation.[3-7] Thus far, few publication have applied text-mining to examine the global relationships between genes and diseases. Perez-Iratxeta et al. automatically examined the GO (Gene Ontology) annotation of genes and their predicted chromosomal locations in order to identify genes linked to inherited disorders.[8]

To obtain a more global understanding of disease development, it would be valuable to incorporate information regarding all possible gene-disease relationships, including biochemical, physiological, pharmacological, epidemiological, as well as genetic. This information would enable comprehensive comparisons between large experimental datasets and existing knowledge in the literature. This would accomplish two things. First, it would serve to validate experiments by demonstrating that known responses occur as predicted. Second, it would rapidly highlight which genes are corroborated by the literature and which genes are novel in a given context. We have utilized a computational approach to literature mining to produce a

* To whom correspondence should be addressed: jlabaer@hms.harvard.edu.

10.1021/pr0340227 CCC: $25.00

Published on Web 05/13/2003

comprehensive set of gene-disease relationships. In addition, we have developed a novel approach to assess the strength of each association based on the frequency of citation and co-citation. We applied this tool to help interpret the data from a large micro-array gene expression experiment comparing normal and cancerous breast tissue.

Methods

MedGene Database. MedGene is a relational database, storing disease and gene information from NCBI, text mining results, statistical scores, and hyperlinks to the primary literature. MedGene has a web-based user interface for users to query the database (http://hipseq.med.harvard.edu/MedGene/).

Text Mining Algorithms. MeSH files were downloaded from the MeSH web site at NLM (Nation Library of Medicine) (http://www.nlm.nih.gov/mesh/meshhome.html) and human disease categories were selected. LocusLink files were downloaded from the LocusLink web site at NCBI (http://www.ncbi.nlh.gov/LocusLink/). Official/preferred gene symbol, official/preferred gene name, and gene alternative symbols and names, all relevant annotations and URLs for each LocusLink record, were collected. Gene search terms were used for literature searching and included all qualified gene names, gene symbols, and gene family terms. Primary gene keys, predominantly qualified gene family terms and gene official/preferred symbols, were used to index Medline records. If the official/preferred gene symbols did not meet the standards to be an index, then qualified gene official/preferred names were used. A local copy of Medline records (up to July, 2002) was pre-selected.

A JAVA module examined the MeSH terms and then indexed each Medline record with the appropriate disease terms. A separate JAVA module was used to examine the titles and abstracts for gene search terms and then to index the gene-related Medline records with the relevant primary gene key(s).

Statistical Methods. For every gene and disease pair, we counted records that were indexed for both gene and disease (double positive hits), for disease only (disease single hits), for gene only (gene single hits), and for neither gene nor disease (double negative hits) to generate a 2 × 2 contingency table. On the basis of the contingency table-framework, we applied different statistical methods to estimate the strength of gene-disease relationships and evaluated the results. These methods included chi-square analysis, Fisher's exact probabilities, relative risk of gene, and relative risk of disease[15] (http://hipseq.med.harvard.edu/MedGene/). In addition, we computed the "product of frequency", which is the product of the proportion of disease/gene double hits to disease single hits and the proportion of disease/gene double hits to gene single hits. To obtain a normal distribution, we transformed all the statistical scores using the natural logarithm. We selected the log of the product of frequency (LPF) to validate MedGene and to use for the analysis with the micro-array data. Spearman rank-correlation coefficients were used to assess the linear relationship between LPF and micro-array fold change in expression level.

Global Analysis. Diseases with at least 50 related genes were selected for clustering analysis, and the LPF scores were normalized with total score for each disease. Hierarchical clustering was done with the "Cluster" software and the clustering result was visualized using "TreeViewer" (http://rana.lbl.gov/EisenSoftware.htm).

Breast Tissue Micro-Arrays. Eighty-nine breast cancer samples (79% ER-positive) and 7 normal breast tissue samples were selected from the Harvard Breast SPORE frozen tissue repository and were representative of the spectrum of histological types, grades, and hormone receptor immuno-phenotypes of breast cancer. Biotinylated cRNA, generated from the total RNA extracted from the bulk tumor, was hybridized to Affymetrix U95A oligo-nucleotide micro-arrays. These micro-arrays consist of 12 400 probes, which represent approximately 9000 genes. Raw expression values were obtained using GENECHIP software from Affymetrix, and then further analyzed using the DNA-Chip Analyzer (dChip) custom software.

Results

Automated Indexing of Medline Records by Disease and Gene. To study the gene-disease associations in the literature, we first compiled complete lists for human diseases and human genes. To index all Medline records that were relevant to human diseases, the Medical Subject Heading (MeSH) index of Medline records was utilized. MeSH is a controlled medical vocabulary from the National Library of Medicine and consists of a set of terms or subject headings that are arranged in both an alphabetic and an hierarchical structure. Medline records are reviewed manually and MeSH terms are added to each with software assistance.[9,10] Twenty-three human disease category headings along with all of their child terms (see the Supporting Information, Supplemental Table 1, or visit http://hipseq.med.harvard.edu/MedGene/publication/s_Table 1.html) were selected from the 2002 MeSH index creating a list of 4033 human diseases.

No index comparable to the MeSH index exists for genes, and thus, it was necessary to apply a string search algorithm for gene names or symbols found in Medline text. A complete list of genes, gene names, gene symbols, and frequently used synonyms were collected from the LocusLink database at NCBI,[11,12] which contains 53 259 independent records keyed by an official gene symbol or name (June 18th, 2002). For the purposes of this study, no distinction was made between genes and their gene products. Authors often use the same name for both, differentiating the two only by the use of italics, if at all. For the intended use of this study, this lack of distinction is unlikely to have a large effect and may in fact be beneficial.

Initial attempts to search the literature using these lists revealed several sources of false positives and false negatives (Table 1). False positives primarily arose when the searched term had other meanings, whereas false negatives arose from syntax discrepancies necessitating the development of filters to reduce these errors. The syntax issues were readily handled by including alternate syntax forms in the search terms. The false positive cases, caused by duplicative and unrelated meanings for the terms, were more difficult to manage. Where possible, case sensitive string mapping reduced inappropriate citations. In many cases, however, this was not sufficient and the terms had to be eliminated entirely, thereby reducing the false positive rate but unavoidably under-representing some genes.

For the purposes of data tracking, a primary gene key was selected to represent all synonyms that correspond to each gene. Medline records were indexed with a primary gene key when any synonym for that key was found in the title or abstract. Case-insensitive string mapping was used for all searches except as noted above. No additional weight was

Table 1. Systematic Sources of False Positives and False Negatives in Unfiltered Data[a]

source of error	error type	example	filter solution
gene symbol/name is not unique	false positive	*MAG*—myelin associated glycoprotein *MAG*—malignancy-associated protein	eliminate this term
gene symbol is unrelated abbreviation	false positive	*PA*—pallid homologue (mouse), pallidin (also abbrev. for Pennsylvania)	eliminate this term
gene symbol/name has language meaning	false positive	*WAS*—Wiskott–Aldrich Syndrome (also the word "was")	case-sensitive string search
nonstandard syntax	false negative	*BAG-1* instead of *BAG1*	add dash term
unofficial gene name/symbol	false negative	*P53* instead of *TP53*	add all gene nicknames
nonspecified gene name	false negative	estrogen receptor instead of Estrogen receptor 1	add family stem term

[a] In preliminary studies, Medline was searched for co-occurrence of genes and diseases and the resulting output was evaluated to identify error sources that were amenable to global filters. Each error source is categorized by the type of error it causes: false positives are suggested relationships that are not real and false negatives are real relationships that are underrepresented. The filter solutions used are indicated. Note that in some cases, the filter solution itself introduces error. In general, error rates maximized sensitivity, even at the expense of specificity if needed.

added for multiple occurrences of a term or the co-occurrence of multiple synonyms for the same gene key.

Medline records were searched with all qualified gene identifiers, such as the official/preferred gene symbol, the official/preferred gene name, all gene nicknames and all syntax variants. In situations where there are several members of a gene family or splice variants, some authors prefer to use a shortened gene family name, e.g., estrogen receptor instead of estrogen receptor 1 (*ESR1*), creating a source of false negatives. For this reason, gene family stem terms were created for all genes that have an alpha or numerical suffix (e.g., *IL2RA*, *TGFβ*, *ESR1*, etc.) and then used to search the literature. The family stem terms were handled separately from the specific gene names so that it would be clear when linkages were made to the gene family versus a specific member in that family.

To improve performance and accuracy, some pre-selection was applied to the records that were scanned. First, review articles were eliminated to avoid redundant treatment of citations. Second, non-English journals were removed because the natural language filters were only relevant to English publications. Finally, journals unlikely to contain primary data about gene-disease relationships were also removed (e.g., *Int. J. Health Educ.*, *Bedside Nurse*, and *J. Health Econ.*). Together, these filters reduced the 12 198 221 Medline publications (July 2002) by 37%.

Ranking the Relative Strengths of Gene-Disease Associations. In total, there were 618 708 gene-disease co-citations, in which 16% (8297) of all studied genes had been associated to a disease and 96% (3875) of all diseases had been associated to at least one gene. To rank the relative strengths of gene disease relationships, we tested several different statistical methods and examined the results. With the exception of the relative risk estimates, the methods provided similar results with respect to the rank order of the gene-disease association strengths. However, after comparing the results to other databases and after consulting disease experts, the log of the product of frequency (LPF) was selected for further analysis because it gave the best results overall.

Validation of MedGene. In developing this tool, it was important to minimize the number of missed genes (false negatives) and miscalled genes (false positives). However, in situations when these goals were in conflict, inclusiveness was prioritized. To determine the false negative rate in MedGene, breast cancer was used as a test case because it was associated with more genes than any other human disease and because

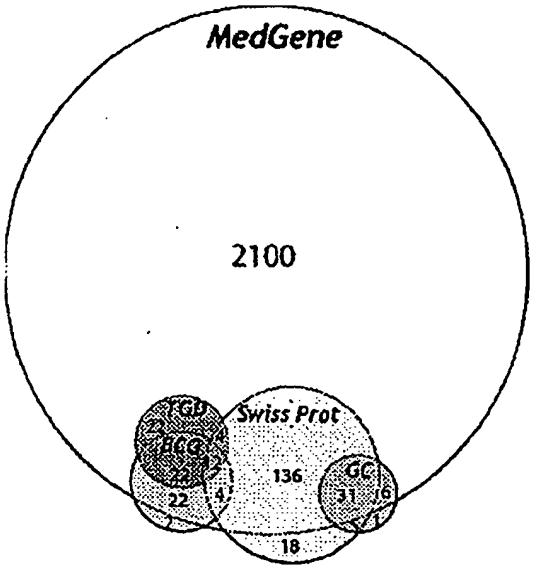


Figure 1. Estimation of the false negative rate by comparison with hand-curated databases. The breast cancer-related genes identified by MedGene were compared with those listed in several other databases including the Tumor Gene Database (TGD),[2] the Breast Cancer Gene Database(BCG),[1] GeneCards (GC)[17] and Swissprot.[18] Genes were considered false negatives if they were represented in at least one of these other databases and not in MedGene and their link to breast cancer was supported by at least one literature reference. All literature references were verified by manual review to confirm their validity. The number of genes in each database or shared by more than one database is indicated. The false negative rate was calculated by genes missed at MedGene (26)/total number of nonoverlapping genes in other databases (285).

there were several public databases that link genes to breast cancer. We compared the list of breast cancer-related genes from MedGene to these databases, illustrated in Figure 1. Among the 285 distinct breast cancer-related genes that were supported by at least one literature citation in these hand-curated databases, 26 were absent from MedGene, suggesting a false negative rate of approximately 9%. To determine why these were missed, all literature references for these genes (80

papers) were reviewed manually (see the Supporting Information, Supplemental Table 2, or visit http://hipseq.med.harvard.edu/MedGene/publication/s_Table 2.html). Among these papers, most false negatives were caused by nonstandard gene terms or gene terms eliminated by our specificity filters. Few genes were missed because they were only mentioned in review papers (0.4%) or they appeared only in the body of the manuscript but not the abstract or title (1.1%). Of note, MedGene identified approximately 2000 additional breast cancer-related genes not listed in any other database.

To assess the false positive error rate, two complementary approaches were used: a detailed analysis of one disease and a global examination of 1000 diseases. The detailed approach examined the false positive error rate and its sources, whereas the global approach tested whether the overall results made biomedical sense.

Using the LPF, 1467 genes related to prostate cancer were assembled in rank order. We then retrieved approximately 300 Medline records each for the highest ranked 100 and the lowest ranked 200 genes and manually reviewed the titles and abstracts to determine the verity of the association. Nearly 80% of the highest ranked 100 genes fell into one of the five categories that reflect meaningful gene-disease relationships (see the Supporting Information, Supplemental Table 3, or visit http://hipseq.med.harvard.edu/MedGene/publication/s_Table 3.html). Among the lowest ranked 200 genes, approximately 70% reflected true relationships. Of the 600 records reviewed, there were only two in which the association between the gene and the disease was described as negative. Both were genes with very low scores. In both cases, the authors did not argue the absence of *any* relationship, but rather that a particular feature of the gene or protein was not shown to be related to human prostate cancer.[13,14]

The coincidence of some gene symbols with medical abbreviations, chemical abbreviations and biological abbreviations resulted in most of the false positives (see the Supporting Information, Supplemental Table 4, or visit http://hipseq.med.harvard.edu/MedGene/publication/s_Table 4.html), emphasizing the importance of the filters that were added in the search algorithm (Table 1). Without the filters, the false positive rate more than doubled, and the false negative rate rose dramatically (data not shown). For example, among the papers about breast cancer, there were only 12 Medline records that referred to *ESR1* and 10 to *ESR2*, whereas almost 2000 papers mentioned estrogen receptor without specifying *ESR1* or *ESR2*; this latter group was detected by the family stem term filter.

To further validate these results, a global analysis of the gene-disease relationships described by MedGene was performed. For this experiment, it was reasoned that the more closely related the diseases are to one another, the more they will be related to the same gene sets. Thus, if the relationships defined by MedGene accurately reflected the literature, then an unsupervised hierarchical clustering of the gene data should group diseases in a manner consistent with common medical thinking. Conversely, if the clustered diseases do not make sense biologically or medically, it may reflect excessive false positives, false negatives, or inappropriate scoring of the data.

To execute this experiment, the gene sets and the corresponding LPF values for 1000 randomly selected diseases (each with at least 50 gene relationships) were used as a dataset for clustering the diseases. A review of the results showed that the resulting disease clusters were indeed logical based upon common medical knowledge (see the Supporting Information, Supplemental Figure 1, or visit http://hipseq.med.harvard.edu/MedGene/publication/s_Figure 1.html). For example, in one such cluster shown in Figure 2, diabetes and its complications grouped together and were also closely linked to diseases associated with starvation states.

The number of genes associated with a given disease can be estimated by adjusting the MedGene number up by the false negative rate (~9%) and down by the false positive rate (~26% on average). Using this, the average disease has 103.7 ± 45.3 (mean ± s.d.) genes associated with it, although the range is quite broad with 2359 genes related to breast cancer, 2122 genes related to lung cancer and no genes related to a number of diseases.

Applying MedGene to the Analysis of Large Datasets. Access to a comprehensive summary of the genes linked to human diseases provided an opportunity to analyze data obtained from a high-throughput experiment. We compared the MedGene breast cancer gene list to a gene expression data set generated from a micro-array analysis comparing breast cancer and normal breast tissue samples. Micro-array analysis identified 2286 genes that had greater than a 1-fold difference in mean expression level between breast cancer samples and normal breast samples. Using MedGene, we sorted the 2286 genes into four classes: 555 genes directly linked to breast cancer in the literature by gene term search (first-degree association by gene name); 328 genes directly linked by family term search (first-degree association by family term); 1021 genes linked to breast cancer only through other breast cancer genes (second-degree association); and 505 genes not previously associated with breast cancer. (See the Supporting Information, Supplemental Figure 2, or visit http://hipseq.med.harvard.edu/MedGene/publication/s_Figure 2.html.) Among the 505 previously unrelated genes, 467 were either newly identified genes or genes that had not previously been associated with any disease. Among the remaining 38 genes, 9 had been related to other cancers, specifically esophageal, colon, uterine, skin, and cervix.

To determine whether the genes highlighted by the micro-array analysis were more likely to have been previously linked to breast cancer in the literature, we created a two-dimensional plot of the fold change of expression level between breast cancer and normal tissue versus the literature score (LPF) (Figure 3A). There was a broad spread of expression changes among the genes directly linked to breast cancer ranging from less than 1-fold change (68%) to over 40-fold (0.3%). Notably, the majority of genes with greater than 10-fold expression changes were linked to breast cancer by first-degree association.

Among all 754 genes directly linked to breast cancer in the literature, there was no correlation between LPF and micro-array fold change ($r = 0.018$, p-value = 0.62). However, when we stratified the analysis based on the magnitude of the fold change, we observed an increasing trend in correlation (Figure 3B) suggesting that genes with a more substantial change in expression level were more likely to have a stronger association in the literature. For genes that had 10-fold change or more in expression level, the correlation increased to 0.41 (p-value = 0.05).

When we evaluated the micro-array data separately for ER positive and ER negative tumors, the trend in correlation between fold change and literature score was highly dependent on estrogen receptor status. Interestingly, there was a similar trend in correlation for ER positive tumors, but no trend in correlation for ER negative tumors.

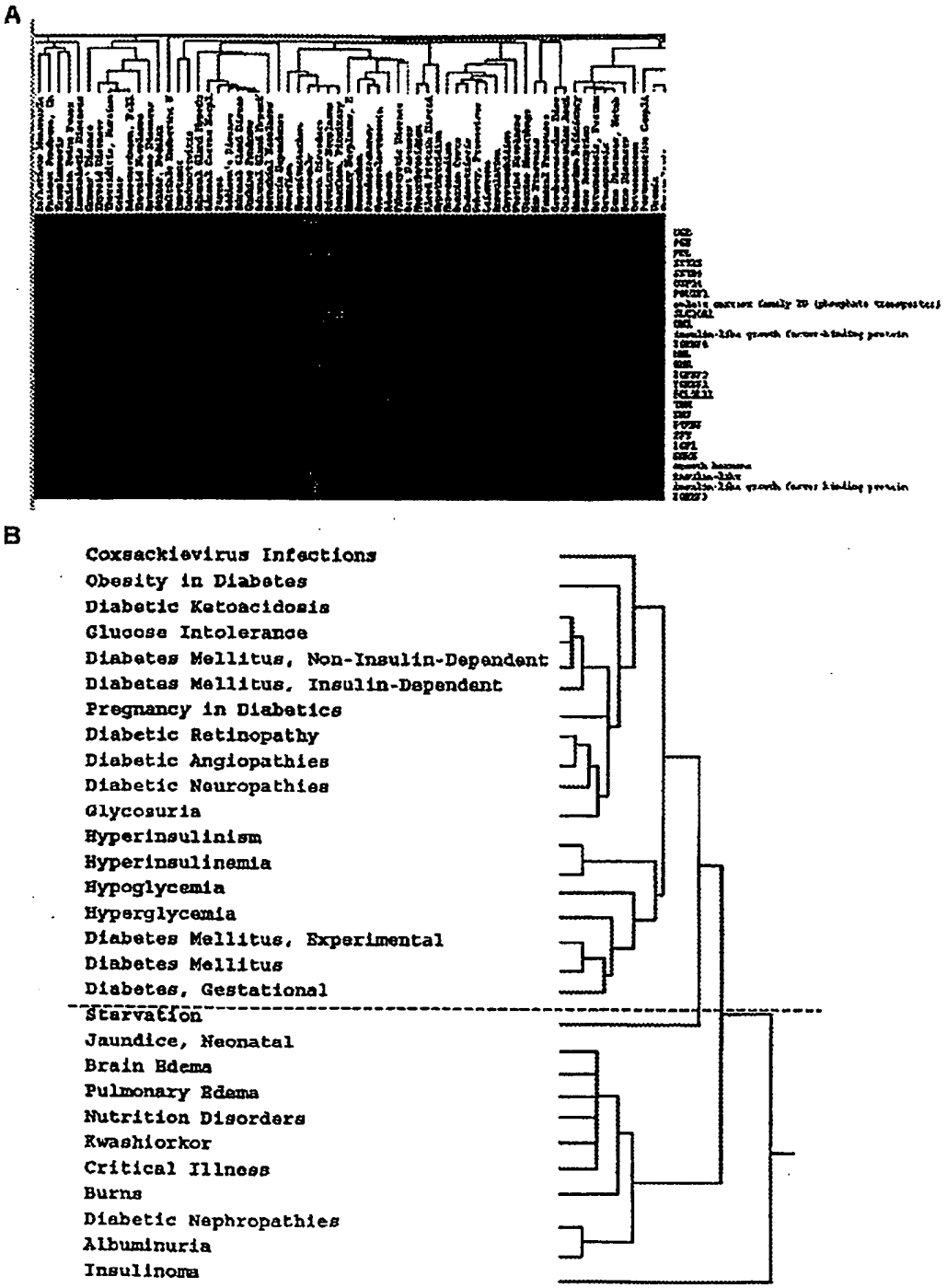


Figure 2. Global validation by clustering analysis. 2(A). The gene sets and the corresponding LPF values for 1000 diseases, each with at least 50 gene relationships, were used in an unsupervised clustering of the diseases based on the gene patterns associated with them. A sample of the data is shown here. 2(B). One of the resulting clusters is shown that corresponds to blood sugar states. Diabetes terms (above the line) and starvation states terms (under the line) clustered together. Within these groups, there is also clustering of diabetic small vessel complications, altered serum chemistries, nutritional disorders, etc.(Supplemental Figure 1: http://hipseq.med.harvard.edu/MedGene/publication/s_Figure 1.html).

Finally, to validate our findings, we computed similar correlations between the breast cancer expression data and LPF scores generated by MedGene for hypertension, a disease unrelated to breast cancer. As expected, we did not observe an increasing trend in correlation for hypertension.

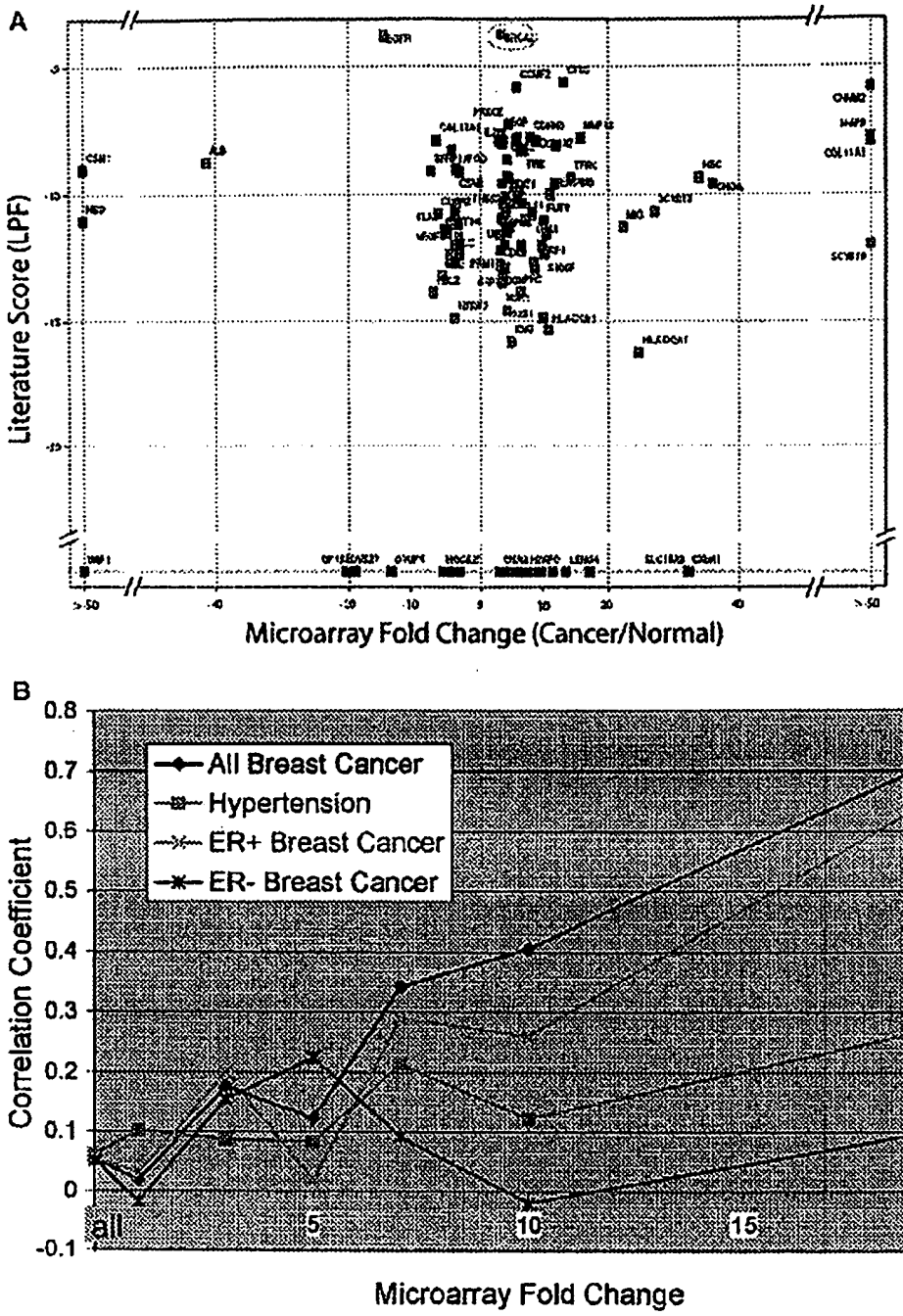


Figure 3. Relationship between literature score and functional data for breast cancer. 3A. The data from an expression analysis of samples for breast tumors and normal breast tissue were analyzed to indicate the fold difference of expression level between breast tumor and normal sample (cutoff ≥ 3-fold change). The fold changes were plotted against the literature score for the same gene set. Green dots represent first-degree association by gene search, blue dots represent first-degree association by family search and red dots represent no-association. Some well-studied genes, such as BRCA2 (pink circle), are not reflected by a substantial difference in expression level. Furthermore, the majority of genes that have no association with breast cancer in the literature had less than 10-fold expression changes (shaded area). 3B. The Spearman rank-correlation coefficients between literature score (LPF) and the fold change of expression level between tumor and normal breast samples (*y*-axis) in relation to the amount of fold change of expression level (*x*-axis). Gene rank lists were generated for breast cancer (blue) and hypertension (pink). Correlations were also computed between the breast cancer gene LPF scores and fold change expression data among estrogen receptor positive tumors only (light blue) and estrogen receptor negative tumors only (purple).

Table 2. Top 25 Genes Related to Selected Human Diseases[a]

breast neoplasms	hypertension	rheumatoid arthritis	bipolar disorder	atherosclerosis
estrogen receptor	*REN*	*RA*	*ERDA1*	apolipoprotein
PGR	*DBP*	*TNFRSF10A*	*SNAP29*	*APOE*
ERBB2	*LEP*	*CRP*	*PFKL*	*LDLR*
BRCA1	*AGT*	*AS*	*DRD2*	*ELN*
BRCA2	*INS*	*ESR1*	*TRH*	*ARG1*
EGFR	kallikrein	*HLA-DRB1*	*IMPA2*	*APOB*
CYP19	*ACE*	*DR1*	*HTR3A*	*APOA1*
TFF1	endothelin	interleukin	*DRD3*	*MSR1*
PSEN2	*S100A6*	*TNF*	*REM*	*LPL*
TP53	*BDK*	*IL6*	*KCNN3*	*PON1*
CES3	*DIANPH*	collagen	*DRD4*	plasminogen activator inhibitor
CEACAM5	*SAR1*	*IL1A*	*HTR2C*	*PLG*
ERBB3	*PIH*	*ACR*	*RELN*	vascular cell adhesion molecule
cyclin	*CD59*	*TNFRSF12*	*DBH*	*ATOH1*
COX5A	*ALB*	*IL2*	*MAOA*	*VWF*
cathepsin	*CYP11B2*	*CHI3L1*	*COMT*	*INS*
ERBB4	*MAT2B*	*IL8*	*HTR2A*	*ARG2*
TRAM	angiotensin receptor	interleukin 1	*SYNJ1*	*ABCA1*
CCND1	*AGTR2*	matrix metalloproteinase	*INPP1*	*OLR1*
EGF	*NPPA*	interferon	*NEDD4L*	collagen
MUC1	*LVM*	*CD68*	*FRA13C*	*MCP*
insulin-like	*DBH*	*IL4*	transducer of *ERBB2*	lipoprotein
BCL2	*NPY*	*IL17*	*BAIAP3*	*APOA2*
mucin	*POMC*	*MMP3*	*ATP1B3*	intercellular adhesion molecule
FGF3	neuropeptide	*SIL*	*DRD5*	*RAB27A*

[a] MedGene results for the top 25 genes associated with breast neoplasms, hypertension, rheumatoid arthritis, bipolar disorder, and atherosclerosis, respectively, ranked by LPF scores. The hyperlink to all the papers co-citing the gene and the disease is available at MedGene website (http://hipseq.med.harvard.edu/MedGene/).

Discussion

The Human Genome Project heralded a new era in biological research where the emphasis on understanding specific pathways has expanded to global studies of genomic organization and biological systems. High-throughput technologies can provide novel insight into comprehensive biological function but also introduces new challenges. The utility of these technologies is limited to the ability to generate, analyze, and interpret large gene lists. MedGene, a relational database derived by mining the information in Medline, was created to address this need. MedGene users can query for a rank-ordered list of human gene-disease relationships (Table 2) for one or more diseases. Each entry is hyperlinked to the original papers supporting each association and to other relevant databases.

MedGene is an innovative extension of previous text mining approaches. Perez-Iratxeta et al. used the GO annotation and their chromosomal locations to predict genes that may contribute to inherited disorders.[6] MedGene takes a broader view and includes all diseases and all possible gene-disease relationships. Furthermore, MedGene utilizes co-citation to indicate a relationship rather than GO annotation, which is limited to the subset of genes that have GO annotation. Our approach is complementary to that taken by Chaussabel and Sher, who used the frequency of co-cited terms to cluster genes into a hierarchy of gene-gene relationships.[5]

A unique aspect of this tool is the ability to assess the relative strengths of gene-disease relationships based on the frequency of both co-citation and single citation. This presupposes that most co-citations describe a positive association, often referred to as publication bias[15] and is supported by our observations that negative associations are rare (Supplemental Table 3: http://hipseq.med.harvard.edu/MedGene/publication/s_Table 3.html). Of course, relationships established by frequency of co-citation do not necessarily represent a true biological link; however, it is strong evidence to support a true relationship.

Another important feature of MedGene is the implementation of software filters that substantially reduced the error rate. We estimate that less than 10% of all associations were missed and at least 70% of even the weakest associations were real. For this study, all of the filters that we applied were general ones, e.g., expanding the list of all gene names to address the different syntax forms used by different journals, eliminating gene names that correspond to common English words, etc. The majority of the remaining search term ambiguities were idiosyncratic and difficult to identify systematically without causing a significant rise in false negatives. Alternative approaches, such as the examination of the nearest neighbor terms, need to be considered to further reduce the false positive rate.

It is not uncommon to see expression changes in microarray experiments as small as 2-fold reported in the literature. Even when these expression changes are statistically significant, it is not always clear if they are biologically meaningful. When comparing expression levels of disease to normal tissue, one expects an enrichment of known disease-related genes to appear in the altered expression group. MedGene provided a unique opportunity to test this notion in the context of existing knowledge on a novel breast cancer micro-array dataset. For genes displaying a 5-fold change or less in tumors compared to normal, there was no evidence of a correlation between altered gene expression and a known role in the disease. This

Table 3. Genes with Large Expression Changes in ER− but Not in ER+ Breast Tumors

gene symbol	fold change (ER+)	fold change (ER−)
KRTHB1	1.0	610.8
BRS3	1.2	89.4
DKK1	1.2	69.8
ZIC1	1.9	59.6
TLR1	1.0	38.5
KIAA0680	2.6	33.2
CDKN3	1.0	30.6
EBI2	4.0	27.9
GZMB	3.8	21.9
STK18	4.7	18.6
GPR49	1.0	14.6
MYO10	1.6	14.4
LAD1	−1.0	13.5
POLE2	4.2	13.0
HMG4	4.4	12.9
BCL2L11	−1.2	12.3
LRP8	2.9	12.2
CCNB2	1.0	11.8
CCNE2	4.0	11.6
FGB	−4.3	11.1
KNSL6	2.9	10.9
H1F5	3.0	10.2
SERPINH2	4.6	10.2
YAP1	1.0	10.0
LPHB	−1.3	−10.4
TCEA2	−1.1	−10.8
TFF1	1.3	−11.4
COL17A1	−4.1	−15.7
POP5	1.1	−16.2
BPAG1	−4.6	−22.3
PDZK1	−1.1	−36.8
VEGFC	−2.8	−51.5
MUC6	−1.4	−64.9
SERPINA5	−1.0	−83.1
MEIS1	−1.6	−85.9
CA12	2.4	−150.3

Table 3. MedGene identified a set of relatively understudied, yet highly expressed genes in ER negative, but not ER positive breast tumors. All of these genes have either never been co-cited with breast cancer or have a weak association except those marked with an *.

reflects the many genes whose role in breast cancer may not involve large changes in expression in sporadic tumors (e.g., *BRCA1* and *BRCA2*) and genes whose modest changes in expression may be unrelated to the disease. Strikingly, among genes with a 10-fold change or more in expression level, there was a strong and significant correlation between expression level and a published role in the disease, providing the first global validation of the micro-array approach to identifying disease-specific genes.

The results derived from MedGene have two implications. First, a careful hunt for corroborating evidence of a role in breast cancer should precede any further study of genes with less than 5-fold expression level changes. Second, any genes with 10-fold changes or more are likely to be related to breast cancer and warrant attention. It is likely that this threshold will change depending on the disease as well as the experiment.

Interestingly, the observed correlation was only found among ER-positive tumors, not ER-negative. This may reflect a bias in the literature to study the more prevalent type of tumor in the population. Furthermore, this emphasizes that caution must be taken when interpreting experiments that may contain subpopulations that behave very differently. The MedGene approach identified a set of relatively understudied, yet highly expressed genes in ER-negative tumors that are worthy of further examination (Table 3).

In conclusion, we have developed an automated method of summarizing and organizing the vast biomedical literature. To our knowledge, the resulting database is the most comprehensive and accurate of its kind. By generating a score that reflects the strength of the association, it provides an important tool for the rapid and flexible analysis of large datasets from various high-throughput screening experiments. Furthermore, it can be used for selecting subsets of genes for functional studies, for building disease-specific arrays, for looking at genes common to multiple diseases and various other high-throughput applications. In the future, it will be possible to enhance the utility of the MedGene database by building links between genes and other MeSH terms as well as other biological processes and concepts, such as cell division and responses to small molecules.

Acknowledgment. We would like to thank P. Braun, L. Garraway, J. Pearlberg, and other members of our institute for helpful discussion. Many thanks to the NLM (National Library of Medicine) for licensing of MEDLINE and the annotation effort of adding MeSH indexes for MEDLINE abstracts. This work was funded by grants from the Breast Cancer Research Foundation and an NHLBI PGA Grant (Vol HL66582-02).

Supporting Information Available: Twenty-three human disease category headings along with all of their child terms selected from the 2002 MeSH index (Supplemental Table 1); analysis of the causes of false negatives in MedGene (Supplemental Table); meaningful gene-disease relationships found in MedGene (Supplemental Table 3); causes for incorrect assignment of gene indexes (Supplemental Table 4); a review of the results, showing that the resulting disease clusters were indeed logical (Supplemental Figure 1); and a review of the results showing that among the 505 previously unrelated genes, 467 were either newly identified genes or genes that had not previously been associated with any disease (Supplemental Figure 2). This material is available free of charge via the Internet at http://pubs.acs.org and at the web sites mentioned in the text.

References

(1) Baasiri, R. A.; Glasser, S. R.; Steffen, D. L.; Wheeler, D. A. *Oncogene* **1999**, *18*, 7958−7965.
(2) Steffen, D. L.; Levine, A. E.; Yarus, S.; Baasiri, R. A.; Wheeler, D. A. *Bioinformatics* **2000**, *16*, 639−649.
(3) Marcotte, E. M.; Xenarios, I.; Eisenberg, D. *Bioinformatics* **2001**, *17*, 359−363.
(4) Ono, T.; Hishigaki, H.; Tanigami, A.; Takagi, T. *Bioinformatics* **2001**, *17*, 155−161.
(5) Jenssen, T. K.; Laegreid, A.; Komorowski, J.; Hovig, E. *Nat. Genet.* **2001**, *28*, 21−28.
(6) Chaussabel, D.; Sher, A. *Genome Biol.* **2002**, *3*, RESEARCH0055.
(7) Gibbons, F. D.; Roth, F. P. *Genome Res.* **2002**, *12*, 1574−1581.
(8) Perez-Iratxeta, C.; Bork, P.; Andrade, M. A. *Nat. Genet.* **2002**, *31*, 316−319.
(9) Funk, M. E.; Reid, C. A. *Bull. Med. Libr. Assoc.* **1983**, *71*, 176−183.
(10) Humphrey, S. M.; Miller, N. E. *J. Am. Soc. Inf. Sci.* **1987**, *38*, 184−196.
(11) Maglott, D. R.; Katz, K. S.; Sicotte, H.; Pruitt, K. D. *Nucleic Acids Res.* **2000**, *28*, 126−128.
(12) Pruitt, K. D.; Maglott, D. R. *Nucleic Acids Res.* **2001**, *29*, 137−140.
(13) Wadelius, M.; Andersson, A. O.; Johansson, J. E.; Wadelius, C.; Rane, E. *Pharmacogenetics* **1999**, *9*, 333−340.
(14) Adam, R. M.; Borer, J. G.; Williams, J.; Eastham, J. A.; Loughlin, K. R.; Freeman, M. R. *Endocrinology* **1999**, *140*, 5866−5875.
(15) Montori, V. M.; Smieja, M.; Guyatt, G. H. *Mayo Clin. Proc.* **2000**, *75*, 1284−1288.
(16) Denenberg, V. H. *Statistics Experimental Design for Behavioral and Biological Researchers*; Wiley-Liss: New York, 1976.
(17) Rebhan, M.; Chalifa-Caspi, V.; Prilusky, J.; Lancet, D. *Trends Genet.* **1997**, *13*, 163.
(18) Bairoch, A.; Apweiler, R. *Nucleic Acids Res.* **2000**, *28*, 45−48.

PR0340227

mRNA Differential display: application in the discovery of novel pharmacological targets

Xinkang Wang, Robert R. Ruffolo, Jr and Giora Z. Feuerstein

The number of genes in the human genome is estimated at 50000–100000. However, only a fraction of these genes are expressed in any given cell. Moreover, the level of gene expression in cells may vary with time, physiological conditions and disease states. This differential gene expression is generally reflected by the different number of mRNA species expressed in a given cell (~15000 individual mRNA species per cell) at any time point, and changes in relative mRNA levels may have important implications in the development of pathological processes. Therefore, discovery of differentially expressed genes is essential for the understanding of the molecular mechanisms involved in normal and pathological states, as well as providing new insights for discovery of new molecular targets for pharmacological manipulation and drug development. Hence, a number of techniques have been developed to identify genes (with known or unknown sequences and functions) that are differentially expressed in disease states. For example, northern hybridization, RNase protection assay, quantitative reverse transcription and polymerase chain reaction (RT-PCR) have been successfully utilized to identify discordantly expressed known genes. Other techniques, such as differential hybridization and subtractive library screening, have been used successfully for the discovery of differentially expressed genes with known and/or unknown sequences. In the differential hybridization method, a cDNA library is first prepared and then screened using probes that are made from two different sources, for example normal and diseased tissues. Subtractive library screening is carried out on the basis of the construction of a subtracted cDNA library from different RNA sources, for example normal and diseased cells, of which theidentical mRNA species have been removed using hybridization methods. Although these two techniques have proved to be useful in the discovery of differentially expressed genes, they are technically difficult and labour intensive, and require large amounts of mRNA (see Box 1).

Recently, a number of PCR-based methods to uncover differentially expressed genes have been developed; these techniques include (1) mRNA differential display[1], (2) RNA fingerprinting[2] and (3) arbitrarily primed PCR (Ref. 3). These PCR-based techniques provide some advantages over the conventional methods and have been used successfully for novel gene discovery. In particular, the mRNA differential display methodology has been adopted by a large number of laboratories as an important additional tool that has applications for both *in vitro* and *in vivo* test systems[1,4–7]. An overall strategic approach using this method for drug discovery is outlined in Fig. 1.

Messenger RNA expression

Messenger RNA is the product of gene expression that encodes for a specific protein. The levels of mRNA in the cell are generally reflected by transcriptional regulation. Following transcription, mRNA is 'matured' by capping the 5'-end, adding the polyadenylation [poly(A)] at the 3'-end, and splicing the intron sequences in eukaryotic cells. Taking advantage of the polyadenylated tail present in most eukaryotic mRNA species, the mRNAs can be reverse-transcribed in the presence of anchored primers complementary to the 3'-end of mRNAs, such as the use of oligo(dT_n) (where n=12–18) primers. In the technique of mRNA differential display, a set of 3'-anchored primers, such as T_{12}MN where M=G, A or C and N=G, A, T or C, are used to prime the reverse transcription reactions.

Methodology: mRNA differential display

The method of mRNA differential display consists of two basic steps: (1) reverse transcription (RT) using a set of 3'-anchored primers, and (2) PCR amplification of cDNA fragments using arbitrary (upstream) primers and anchored (downstream) primers (Fig. 2).

For the RT reaction, total cellular RNA (DNase treated to eliminate the possibility of genomic DNA contamination) is reverse-transcribed to yield the first strand cDNA primed with T_{12}MN oligonucleotides. This RT reaction enables all the mRNA species having a poly(A) tail to be reverse-transcribed. Typically, this RT reaction is divided into four subgroups, each using a different T_{12}MN primer with G, A, T or C at the last base of the 3'-end. Because a large number of mRNA species are present in a cell, the division of subgroups for the RT allows a portion of the mRNA species to be displayed, which will increase the resolution of cDNA species after amplification[1].

Amplification of all the cDNAs is carried out using an upstream arbitrary primer and a downstream anchored primer (identical to the one used for the RT) in the presence of a radioactive nucleotide (Fig. 2). The upstream primer has been optimized to ten bases in length, containing approximately 50% of GC contents[1]. In addition, a relatively low annealing temperature (42°C) is also recommended for the PCR to allow some base mismatches so that a larger number of the amplified mRNA species can be obtained. Using these conditions of amplification, it has been estimated that at

X. Wang, Investigator, R. R. Ruffolo, Jr Vice President and Director, and G. Z. Feuerstein, Cardiovascular Pharmacology Director, Division of Pharmacological Sciences, SmithKline Beecham Pharmaceuticals, King of Prussia, PA 19406, USA.

  0165 – 6147/96/$15.00 PII: S0165-6147(96)40001-3

Box 1. Comparison of mRNA differential display with subtractive library screening for novel gene discovery

mRNA Differential display	Subtractive library screening
• Key technique is based on RT-PCR	• Crucial step is subtractive library construction
• Very sensitive to detect altered gene expression	• Relatively insensitive, especially for those low abundance mRNAs
• Allows multiple comparison, and monitors both upregulated and downregulated genes	• Usually compares only unidirectional change
• Relatively reliable to detect the differentially expressed genes; confirmation by other techniques is required	• Very reliable to detect the altered gene expression
• Rapid to identify a lead probe	• Relatively slow and complicated

least 30–40 upstream primers in combination with the downstream primers will be necessary to amplify every mRNA species present in a given cell[8].

The amplified cDNA fragments are resolved by electrophoresis and subjected to autoradiographic analysis. By taking advantage of mRNA differential display, multiple samples can be amplified and compared in parallel. As such, differences in gene expression, either upregulated or downregulated, can be identified in specific experimental or pathological conditions or along temporal expression patterns. As shown in Figure 3, the differential display analysis was carried out using cellular RNAs isolated from lipopolysaccharide (LPS)-stimulated and -unstimulated rat aortic vessels[9].

Band recovery

Following mRNA differential display, the bands of interest may be recovered by applying the following three steps: the DNA band is (1) excised from the dried sequencing gel, (2) isolated by extraction procedures, and (3) reamplified using the same sets of primers as in the original PCR (Ref. 1). The recovered DNA band can serve as a probe to confirm mRNA expression by means of northern blot analysis, and/or be subcloned into a vector for further analysis.

Confirmation of the differentially expressed genes

Confirmation of gene expression is one of the crucial steps following mRNA differential display, in as much as a large number of false-positive bands may be present on differential display. A variety of methods to reduce false positives have been utilized in different laboratories; the most commonly used method is northern blot analysis. Dot blot, quantitative RT–PCR, RNase protection assays and other methods have also been used.

Using two methods, differential display and northern blot analyses, the significant upregulation of mRNA (LPS-7) in response to LPS stimulation in cultured aortic vessels has been confirmed (Figs 3 and 4; see Ref. 9).

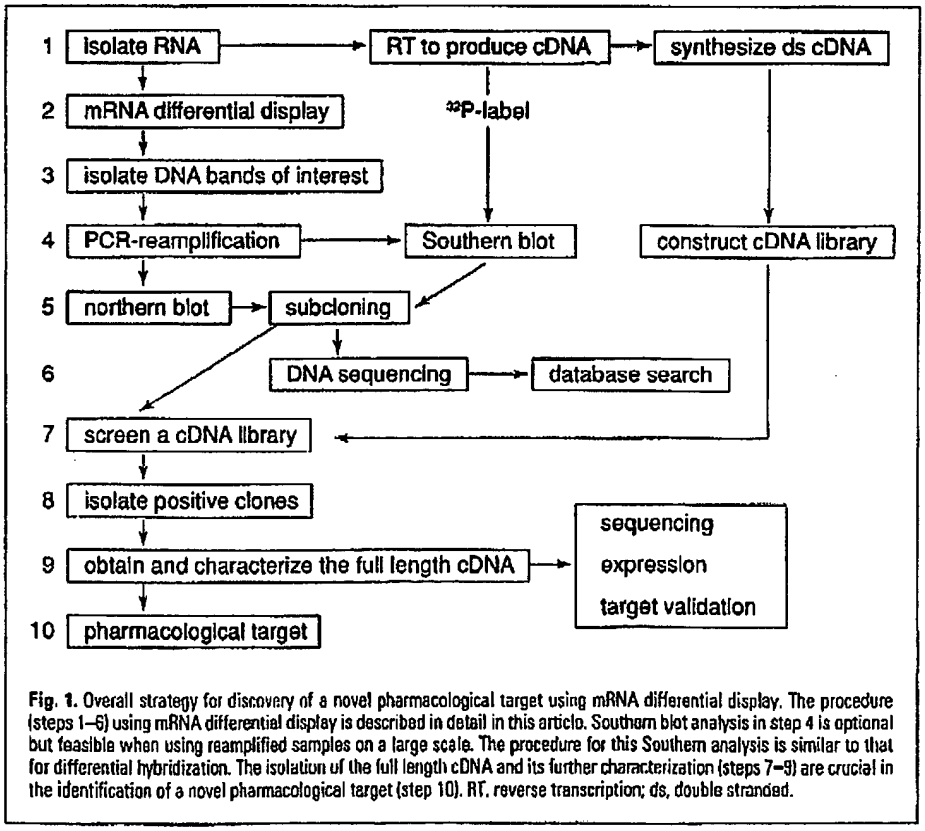


Fig. 1. Overall strategy for discovery of a novel pharmacological target using mRNA differential display. The procedure (steps 1–6) using mRNA differential display is described in detail in this article. Southern blot analysis in step 4 is optional but feasible when using reamplified samples on a large scale. The procedure for this Southern analysis is similar to that for differential hybridization. The isolation of the full length cDNA and its further characterization (steps 7–9) are crucial in the identification of a novel pharmacological target (step 10). RT, reverse transcription; ds, double stranded.

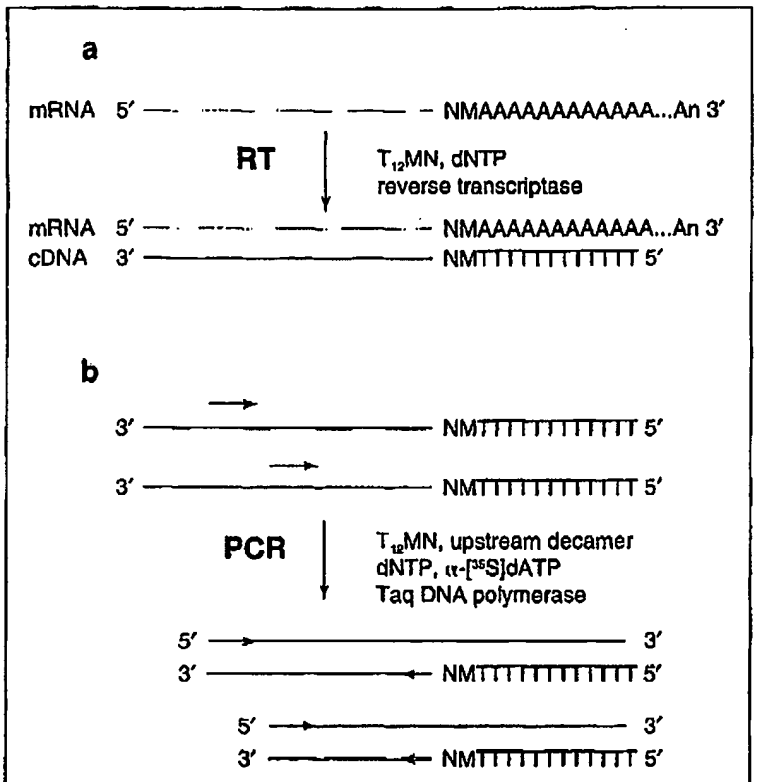


Fig. 2. Outline of the mRNA differential display technique. **a:** reverse transcription (RT) is carried out using primers $T_{12}MN$ (M = G, A or C; N = G, A, T or C) in the presence of dNTP (deoxyribonucleoside triphosphate) and reverse transcriptase. The T_{12} sequence of the primers can anneal to the poly(A) tail of any mRNA species, whereas the last two bases at the 3′-end of the primers (MN) determine the specificity of the RT. **b:** PCR amplification is carried out using the same 3′-anchored primers as in the RT reaction in combination with a particular upstream decamer of random sequence. This reaction is usually labelled with radioactive dNTPs. The decamer will anneal to the cDNAs with the complementary sequences and the region between this sequence and the 3′ anchored position will be amplified.

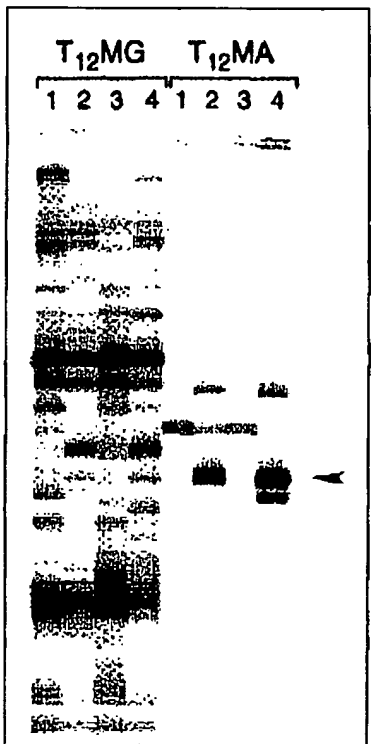


Fig. 3. Differential display of mRNA isolated from cultured rat aortic artery stimulated with LPS (lipopolysaccharide). Specifically, an upstream primer (5′-GACCGCTTGT-3′) in combination with downstream primers, either $T_{12}MG$ (left) or $T_{12}MA$ (right), were used for the amplification. PCR products were resolved in an 8M urea/6% polyacrylamide DNA sequencing gel in the following order: lane 1, unstimulated and lane 2, stimulated aorta from spontaneously hypertensive rats; lane 3, unstimulated and lane 4, stimulated aorta from Wistar–Kyoto rats. The band(s) indicated with an arrowhead (designated as LPS-7) shows a marked induction in response to LPS stimulation on the differential displaying gel.

Identification of the differentially expressed genes

It is fundamental to identify the genes discovered by mRNA differential display. This step relies on the DNA sequencing analysis of the recovered DNA band. Because the primers used for differential display are short and cannot be used successfully for direct sequencing by standard protocols, the differential displayed DNA fragments are typically subcloned into a vector prior to sequencing analysis[1,5]. Recently, direct sequencing of differential display PCR products became feasible (1) on the basis of the use of enlongated primers for direct differential display[10,11] or (2) during the reamplification following original differential display method[9].

Using this sequence information, the identity of the differentially expressed genes can be determined by searching a database, such as GenBank. If the sequence represents an unknown sequence, a cDNA library can be screened using this DNA as a probe in order to obtain the full length cDNA clone.

Advantages of mRNA differential display

Compared with the conventional methods for the discovery of genes with altered expression in disease states, such as differential hybridization and subtractive library screening, the mRNA differential display technique has several advantages (see Box 1): (1) simplicity in all key techniques (primarily RT–PCR); (2) sensitivity due to PCR amplification; (3) versatility in detecting genes that are either upregulated or downregulated under various conditions, and the ability to perform a side-by-side comparison of different samples; (4) rapidity in identifying a probe (a cDNA) and confirming the results (e.g. northern blot); (5) small amounts of RNA required; and (6) reproducibility (the displayed bands in general show at least 60–70% identity for different repeats). These characteristics render this technique increasingly popular for the discovery of novel genes.

Limitations of mRNA differential display

While the differential display technique has significant advantages, some disadvantages in using mRNA differential display must be acknowledged[12]. The major concerns are the high incidence of false positives, and the labour-intensive nature of this procedure for large-scale screening. In addition, the cDNA fragments isolated by this method are typically small, and frequently located in the 3'-untranslated region. Therefore, in order to identify the differentially expressed gene, one may need to screen a cDNA library to isolate the full length cDNA clone. Moreover, in order to observe every differentially expressed gene in the mRNA population, at least 20–25 (and possibly up to 80, see Ref. 13) upstream primers in combination with downstream anchored primers should be used (based upon theoretical calculations)[8]. It is obvious that this technique needs to be refined further in order to be efficiently and widely applied for large-scale searching of altered gene expression in different diseases or under different experimental conditions.

Recently, significant improvements and modifications have been made to the method as originally described[1] in order to overcome some of the existing problems in this technique[14], e.g. (1) emphasis has been placed on the importance of DNA-free RNA samples and multiple displays of samples; this will reduce the frequency of false positives[15]; (2) longer primers are used, e.g. 18–20 mers, as in RNA-fingerprinting[2]; this not only increases the reproducibility of differential display, but also allows direct sequencing after PCR amplification[10,11]; (3) the application of slot blot has been used to evaluate the bands identified after differential display[16], or the use of northern blot for affinity capturing of cDNAs (Ref. 17); these methods reduce the labour-intensive nature of this work for large scale screening. Furthermore, (4) the potential hazardous nature of ^{35}S as a radiolabel for differential display has been noted, and either ^{32}P or ^{33}P have been recommended as alternative labels[18,19].

Concluding remarks

Differential display of mRNA is one of the most flexible and comprehensive methods available for the detection of differentially expressed genes in the cell. Since its initial description, this technique has been established in many laboratories and applied successfully in the identification of genes using *in vitro* and *in vivo* systems. In addition, other strategies aimed at discovering novel genes are emerging, such as methodology of serial analysis of gene expression (SAGE)[20] and representational difference analysis (RDA)[21]. The application of mRNA differential display, and other techniques, for the isolation of novel genes associated with disease processes will no doubt facilitate the discovery of novel therapeutic targets and/or will help to understand the molecular mechanisms of disease. However, this is the first of many steps (Fig. 1) required in the discovery of a novel pharmacological target, especially given that the function of this factor is most likely unknown. Therefore, further action should be taken to characterize the functions of a particular gene of interest, including isolation of full length cDNA, expression of the gene product for functional study and target validation for the importance of this gene in disease processes.

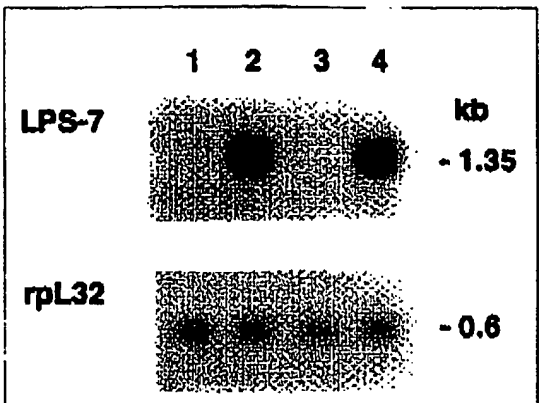


Fig. 4. Northern analysis of LPS-7 mRNA expression in cultured aorta stimulated with LPS. Total cellular RNA (10 μg/lane, loaded in the order: lane 1, unstimulated and lane 2, stimulated aorta from spontaneously hypertensive rats; lane 3, unstimulated and lane 4, stimulated aorta from Wistar–Kyoto rats) was resolved by electrophoresis, transferred to a nylon membrane, and hybridized to LPS-7 and rpl 32 (loading control) cDNA probe sequentially. The rpL32 mRNA expression was relatively constant in the experimental conditions and therefore used for standardizing the samples loaded in each lane.

References

1 Liang, P. and Pardee, A. B. (1992) *Science* 257, 967–971
2 Welsh, J. *et al.* (1992) *Nucleic Acids Res.* 20, 4965–4970
3 Sokolov, B. P. and Prockop, D. J. (1994) *Nucleic Acids Res.* 22, 4009–4015
4 Aiello, L. P., Robinson, G. S., Lin, Y. W., Nishio, Y. and King, G. L. (1994) *Proc. Natl. Acad. Sci. U. S. A.* 91, 6231–6235
5 Zimmermann, J. W. and Schultz, R. M. (1994) *Proc. Natl. Acad. Sci. U. S. A.* 91, 5456–5460
6 Utans, U., Liang, P., Wyner, L. R., Karnovsky, M. J. and Russell, M. E. (1994) *Proc. Natl. Acad. Sci. U. S. A.* 91, 6463–6467
7 Wang, X. K. *et al.* (1995) *Proc. Natl. Acad. Sci. U. S. A.* 92, 11480–11484
8 Bauer, D. *et al.* (1993) *Nucleic Acids Res.* 21, 4272–4280
9 Wang, X. K. and Feuerstein, G. Z. (1995) *BioTechniques* 18, 448–453
10 Zhao, S., Oui, S. J. and Pardee, A. B. (1995) *BioTechniques* 18, 842–850
11 Diachenko, L. B., Ledesma, J., Chenchik, A. A. and Siebert, P. D. (1996) *Biochem. Biophys. Res. Commun.* 219, 824–828
12 Debouck, C. (1995) *Curr. Opin. Biotechnol.* 6, 597–599
13 Liang, P. *et al.* (1995) *Methods Enzymol.* 254, 304–321
14 Liang, P. and Pardee, A. B. (1995) *Curr. Opin. Immunol.* 7, 274–280
15 Liang, P., Averboukh, L. and Pardee, A. B. (1993) *Nucleic Acids Res.* 21, 3269–3275
16 Mou, L., Miller, H., Li, J., Wang, E. and Chalifour, L. (1994) *Biochem. Biophys. Res. Commun.* 199, 564–569
17 Li, F., Barnathan E. S. and Kariko, K. (1994) *Nucleic Acids Res.* 22, 1764–1765
18 Trentmann, S. M., Van der Knaap, E. and Kende, H. (1995) *Science* 267, 1186–1197
19 Tokuyama, Y. and Takeda, J. (1995) *BioTechniques* 18, 424–425
20 Velculescu, V. E., Zhang, L., Vogelstein, B. and Kinzler, K. W. (1995) *Science* 270, 484–487
21 Lisitsyn, N., Lisitsyn, N. and Wigler, M. (1993) *Science* 259, 946–951

Electrophoresis 1998, 19, 1862–1871

Review

Proteome analysis: Biological assay or data archive?

Paul A. Haynes
Steven P. Gygi
Daniel Figeys
Ruedi Aebersold

Department of Molecular Biotechnology, University of Washington, Seattle, WA, USA

In this review we examine the current state of proteome analysis. There are three main issues discussed: why it is necessary to study proteomes; how proteomes can be analyzed with current technology; and how proteome analysis can be used to enhance biological research. We conclude that proteome analysis is an essential tool in the understanding of regulated biological systems. Current technology, while still mostly limited to the more abundant proteins, enables the use of proteome analysis both to establish databases of proteins present, and to perform biological assays involving measurement of multiple variables. We believe that the utility of proteome analysis in future biological research will continue to be enhanced by further improvements in analytical technology.

Contents

1 Introduction

A proteome has been defined as the protein complement expressed by the genome of an organism, or, in multicellular organisms, as the protein complement expressed by a tissue or differentiated cell [1]. In the most common implementation of proteome analysis the proteins extracted from the cell or tissue analyzed are separated by high resolution two-dimensional gel electrophoresis (2-DE), detected in the gel and identified by their amino acid sequence. The ease, sensitivity and speed with which gel-separated proteins can be identified by the use of recently developed mass spectrometric techniques have dramatically increased the interest in proteome technology. One of the most attractive features of such analyses is that complex biological systems can potentially be studied in their entirety, rather than as a multitude of individual components. This makes it far easier to uncover the many complex, and often obscure, relationships between mature gene products in cells. Large-scale proteome characterization projects have been undertaken for a number of different organisms and cell types. Microbial proteome projects currently in progress include, for example: *Saccharomyces cerevisiae* [2], *Salmonella enterica* [3], *Spiroplasma melliferum* [4], *Mycobacterium tuberculosis* [5], *Ochrobactrum anthropi* [6], *Haemophilus influenzae* [7], *Synechocystis spp.* [8], *Escherichia coli* [9], *Rhizobium leguminosarum* [10], and *Dictyostelium discoideum* [11]. Proteome projects underway for tissues of more complex organisms include those for: human bladder squamous cell carcinomas [12], human liver [13], human plasma [13], human keratinocytes [12], human fibroblasts [12], mouse kidney [12], and rat serum [14]. In this manuscript we critically assess the concept of proteome analysis and the technical feasibility of establishing complete proteome maps, and discuss ways in which proteome analysis and biological research intersect.

2 Rationale for proteome analysis

The dramatic growth in both the number of genome projects and the speed with which genome sequences are being determined has generated huge amounts of sequence information, for some species even complete genomic sequences ([15–17]). The description of the state of a biological system by the quantitative measurement of system components has long been a primary objective in molecular biology. With recent technical advances including the development of differential display-PCR [18], cDNA microarray and DNA chip technology [19, 20] and serial analysis of gene expression (SAGE) [21, 22], it is now feasible to establish global and quantitative mRNA expression maps of cells and tissues, in which the sequence of all the genes is known, at a speed and sensitivity which is not matched by current

Correspondence: Professor Ruedi Aebersold, Department of Molecular Biotechnology, University of Washington, Box 357730, Seattle, WA, 98195, USA (Tel: +206-685-4235; Fax: +206-685-6392; E-mail: ruedi@u.washington.edu)

Abbreviations: CID, collision-induced dissociation; MS/MS, tandem mass spectrometry; SAGE, serial analysis of gene expression

Keywords: Proteome / Two-dimensional polyacrylamide gel electrophoresis / Tandem mass spectrometry

0173-0835/98/1111-1862 $17.50+.50/0

protein analysis technology. Given the long-standing paradigm in biology that DNA synthesizes RNA which synthesizes protein, and the ability to rapidly establish global, quantitative mRNA expression maps, the questions which arise are why technically complex proteome projects should be undertaken and what specific types of information could be expected from proteome projects which cannot be obtained from genomic and transcript profiling projects. We see three main reasons for proteome analysis to become an essential component in the comprehensive analysis of biological systems. (i) Protein expression levels are not predictable from the mRNA expression levels, (ii) proteins are dynamically modified and processed in ways which are not necessarily apparent from the gene sequence, and (iii) proteomes are dynamic and reflect the state of a biological system.

2.1 Correlation between mRNA and protein expression levels

Interpretations of quantitative mRNA expression profiles frequently implicitly or explicitly assume that for specific genes the transcript levels are indicative of the levels of protein expression. As part of an ongoing study in our laboratory, we have determined the correlation of expression at the mRNA and protein levels for a population of selected genes in the yeast *Saccharomyces cerevisiae* growing at mid-log phase (S. P. Gygi *et al.*, submitted for publication). mRNA expression levels were calculated from published SAGE frequency tables [22]. Protein expression levels were quantified by metabolic radiolabeling of the yeast proteins, liquid scintillation counting of the protein spots separated by high resolution 2-DE and mass spectrometric identification of the protein(s) migrating to each spot. The selected 80 samples constitute a relatively homogeneous group with respect to predicted half-life and expression level of the protein products. Thus far, we have found a general trend but no strong correlation between protein and transcript levels (Fig. 1). For some genes studied equivalent mRNA transcript levels translated into protein abundances which varied by more than 50-fold. Similarly, equivalent steady-state protein expression levels were maintained by transcript levels varying by as much as 40-fold (S. P. Gygi *et al.*, submitted). These results suggests that even for a population of genes predicted to be relatively homogeneous with respect to protein half-life and gene expression, the protein levels cannot be accurately predicted from the level of the corresponding mRNA transcript.

2.2 Proteins are dynamically modified and processed

In the mature, biologically active form many proteins are post-translationally modified by glycosylation, phosphorylation, prenylation, acylation, ubiquitination or one or more of many other modifications [23] and many proteins are only functional if specifically associated or complexed with other molecules, including DNA, RNA, proteins and organic and inorganic cofactors. Frequently, modifications are dynamic and reversible and may alter the precise three-dimensional structure and the state of activity of a protein. Collectively, the state of modification of the proteins which constitute a biological system

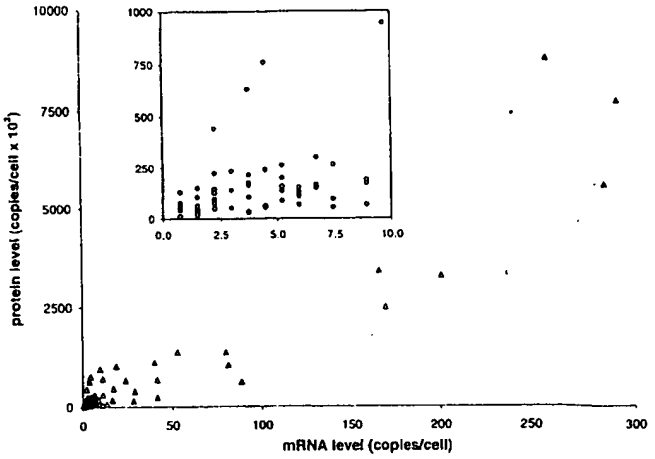


Figure 1. Correlation between mRNA and protein levels in yeast cells. For a selected population of 80 genes, protein levels were measured by ^{35}S-radiolabeling and mRNA levels were calculated from published SAGE tables. Inset: expanded view of the low abundance region. For more experimental details, also see Figs. 5 and 6, (S. P. Gygi *et al.*, submitted).

are important indicators for the state of the system. The type of protein modification and the sites modified at a specific cellular state can usually not be determined from the gene sequence alone.

2.3 Proteomes are dynamic and reflect the state of a biological system

A single genome can give rise to many qualitatively and quantitatively different proteomes. Specific stages of the cell cycle and states of differentiation, responses to growth and nutrient conditions, temperature and stress, and pathological conditions represent cellular states which are characterized by significantly different proteomes. The proteome, in principle, also reflects events that are under translational and post-translational control. It is therefore expected that proteomics will be able to provide the most precise and detailed molecular description of the state of a cell or tissue, provided that the external conditions defining the state are carefully determined. In answer to the question of whether the study of proteomes is necessary for the analysis of biomolecular systems, it is evident that the analysis of mature protein products in cells is essential as there are numerous levels of control of protein synthesis, degradation, processing and modification, which are only apparent by direct protein analysis.

3 Description and assessment of current proteome analysis technology

3.1 Technical requirements of proteome technology

In biological systems the level of expression as well as the states of modification, processing and macro-molecular association of proteins are controlled and modulated depending on the state of the system. Comprehensive analysis of the identity, quantity and state of modification of proteins therefore requires the detection and

quantitation of the proteins which constitute the system, and analysis of differentially processed forms. There are a number of inherent difficulties in protein analysis which complicate these tasks. First, proteins cannot be amplified. It is possible to produce large amounts of a particular protein by over-expression in specific cell systems. However, since many proteins are dynamically post-translationally modified, they cannot be easily amplified in the form in which they finally function in the biological system. It is frequently difficult to purify from the native source sufficient amounts of a protein for analysis. From a technological point of view this translates into the need for high sensitivity analytical techniques. Second, many proteins are modified and processed post-translationally. Therefore, in addition to the protein identity, the structural basis for differentially modified isoforms also needs to be determined. The distribution of a constant amount of protein over several differentially modified isoforms further reduces the amount of each species available for analysis. The complexity and dynamics of post-translational protein editing thus significantly complicates proteome studies. Third, proteins vary dramatically with respect to their solubility in commonly used solvents. There are few, if any, solvent conditions in which all proteins are soluble and which are also compatible with protein analysis. This makes the development of protein purification methods particularly difficult since both protein purification and solubility have to be achieved under the same conditions. Detergents, in particular sodium dodecyl sulfate (SDS), are frequently added to aqueous solvents to maintain protein solubility. The compatibility with SDS is a big advantage of SDS polyacrylamide gel electrophoresis (SDS-PAGE) over other protein separation techniques. Thus, SDS-PAGE and two-dimensional gel electrophoresis, which also uses SDS and other detergents, are the most general and preferred methods for the purification of small amounts of proteins, provided that activity does not necessarily need to be maintained. Lastly, the number of proteins in a given cell system is typically in the thousands. Any attempt to identify and categorize all of these must use methods which are as rapid as possible to allow completion of the project within a reasonable time frame. Therefore, a successful, general proteomics technology requires high sensitivity, high throughput, the ability to differentiate differentially modified proteins, and the ability to quantitatively display and analyze all the proteins present in a sample.

3.2 2-D electrophoresis – mass spectrometry: a common implementation of proteome analysis

The most common currently used implementation of proteome analysis technology is based on the separation of proteins by two-dimensional (IEF/SDS-PAGE) gel electrophoresis and their subsequent identification and analysis by mass spectrometry (MS) or tandem mass spectrometry (MS/MS). In 2-DE, proteins are first separated by isoelectric focusing (IEF) and then by SDS-PAGE, in the second, perpendicular dimension. Separated proteins are visualized at high sensitivity by staining or autoradiography, producing two-dimensional arrays of proteins. 2-DE gels are, at present, the most commonly used means of global display of proteins in complex samples. The separation of thousands of proteins has been achieved in a single gel [24, 25] and differentially modified proteins are frequently separated. Due to the compatibility of 2-DE with high concentrations of detergents, protein denaturants and other additives promoting protein solubility, the technique is widely used.

The second step of this type of proteome analysis is the identification and analysis of separated proteins. Individual proteins from polyacrylamide gels have traditionally been identified using *N*-terminal sequencing [26, 27], internal peptide sequencing [28, 29], immunoblotting or comigration with known proteins [30]. The recent dramatic growth of large-scale genomic and expressed sequence tag (EST) sequence databases has resulted in a fundamental change in the way proteins are identified by their amino acid sequence. Rather than by the traditional methods described above, protein sequences are now frequently determined by correlating mass spectral or tandem mass spectral data of peptides derived from proteins, with the information contained in sequence databases [31–33].

There are a number of alternative approaches to proteome analysis currently under development. There is considerable interest in developing a proteome analysis stragegy which bypasses 2-DE altogether, because it is considered a relatively slow and tedious process, and because of perceived difficulties in extracting proteins from the gel matrix for analysis. However, 2-DE as a starting point for proteome analysis has many advantages compared to other techniques available today. The most significant strengths of the 2-DE-MS approach include the relatively uniform behavior of proteins in gels, the ability to quantify spots and the high resolution and simultaneous display of hundreds to thousands of proteins within a reasonable time frame.

A schematic diagram of a typical procedure of the identification of gel-separated proteins is shown in Fig. 2. Protein spots detected in the gel are enzymatically or chemically fragmented and the peptide fragments are isolated for analysis, as already indicated, most frequently by MS or MS/MS. There are numerous protocols for the generation of peptide fragments from gel-separated proteins. They can be grouped into two categories, digestion in the gel slice [28, 34] or digestion after electrotransfer out of the gel onto a suitable membrane ([29, 35–37] and reviewed in [38]). In most instances either technique is applicable and yields good results. The analysis of MS or MS/MS data is an important step in the whole process because MS instruments can generate an enormous amount of information which cannot easily be managed manually. Recently, a number of groups have developed software systems dedicated to the use of peptide MS and MS/MS spectra for the identification of proteins. Proteins are identified by correlating the information contained in the MS spectra of protein digests or MS/MS spectra of individual peptides with data contained in DNA or protein sequence databases.

The systems we are currently using in our laboratory are based on the separation of the peptides contained in protein digests by narrow bore or capillary liquid chromatog-

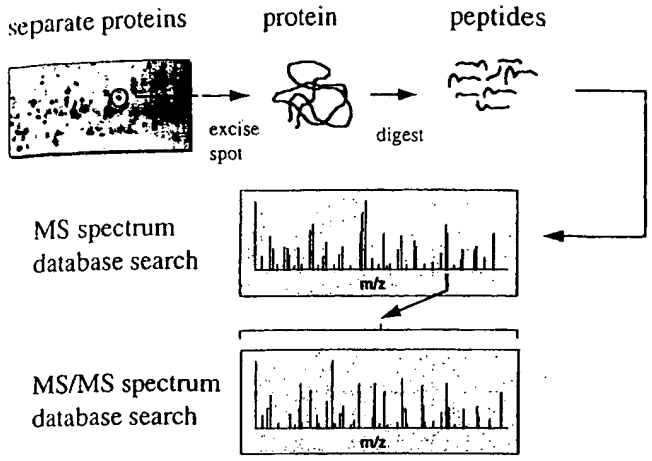


Figure 2. Schematic diagram of a procedure for identification of gel-separated proteins. Peptides can either be separated by a technique such as LC or CE, or infused as a mixture and sorted in the MS. Database searching can either be performed on peptide masses from an MS spectrum, peptide fragment masses from CID spectra of peptides, or a combination of both.

raphy [39, 40] or capillary electrophoresis [41], the analysis of the separated peptides by electrospray ionization (ESI) MS/MS, and the correlation of the generated peptide spectra with sequence databases using the SEQUEST program developed at the University of Washington [32, 33]. The system automatically performs the following operations: a particular peptide ion characterized by its mass-to-charge ratio is selected in the MS out of all the peptide ions present in the system at a particular time; the selected peptide ion is collided in a collision cell with argon (collision-induced dissociation, CID) and the masses of the resulting fragment ions are determined in the second sector of the tandem MS; this experimentally determined CID spectrum is then correlated with the CID spectra predicted from all the peptides in a sequence database which have essentially the same mass as the peptide selected for CID; this correlation matches the isolated peptide with a sequence segment in a database and thus identifies the protein from which the peptide was derived. There are a number of alternative programs which use peptide CID spectra for protein identification, but we use the SEQUEST system because it is currently the most highly automated program and has proven to be successful, versatile and robust.

3.3 Protein identification by LC-MS/MS, capillary LC-MS/MS and CE-MS/MS

It has been demonstrated repeatedly that MS has a very high intrinsic sensitivity. For the routine analysis of gel-separated proteins at high sensitivity, the most significant challenge is the handling of small amounts of sample. The crux of the problem is the extraction and transferal of peptide mixtures generated by the digestion of low nanogram amounts of protein, from gels into the MS/MS system without significant loss of sample or introduction of unwanted contaminants. We employ three different systems for introducing gel-purified samples into an MS, depending on the level of sensitivity required. As an approximate guideline, for samples containing tens of picomoles of peptides, LC-MS/MS is most appropriate; for samples containing low picomole amounts to high femtomole amounts we use capillary LC-MS/MS; and for samples containing femtomoles or less, CE-MS/MS is the method of choice.

3.3.1 LC-MS/MS

The coupling of an MS to an HPLC system using a 0.5 mm diameter or bigger reverse phase (RP) column has been described in detail [42]. This system has several advantages if a large number of samples are to be analyzed and all are available in sufficient quantity. The LC-MS and database searching program can be run in a fully automated mode using an autosampler, thus maximizing sample throughput and minimizing the need for operator interference. The relatively large column is tolerant of high levels of impurities from either gel preparation or sample matrix. Lastly, if configured with a flow-splitter and micro-sprayer [40], analyses can be performed on a small fraction of the sample (less than 5%) while the remainder of the sample is recovered in very pure solvents. This latter feature is particularly useful when an orthogonal technique is also used to analyze peptide fractions, such as scintillation of an introduced radiolabel, and this data can be correlated with peptides identified by CID spectra.

3.3.2 Capillary LC-MS

An increase of sensitivity of approximately tenfold can be achieved by using a capillary LC system with a 100 μm ID column rather than a 0.5 mm ID column as referred to above. Since very low flow rates are required for such columns, most reports have used a precolumn flow splitting system for producing solvent gradients. We have recently desribed the design and construction of a novel gradient mixing system which enables the formation of reproducible gradients at very low flow rates (low nL/min) without the need for flow splitting (A. Ducret *et al.*, submitted for publication). Using this capillary LC-MS/MS system we were able to identify gel-separated proteins if low picomole to high femtomole amounts were loaded onto the gel [40]. This system is as yet not automated and, like all capillary LC systems, is prone to blockage of the columns by microparticulates when analyzing gel-separated proteins.

3.3.3 CE-MS/MS

The highest level of sensitivity for analyzing gel-separated proteins can be achieved by using capillary electrophoresis – mass spectrometry (CE-MS). We have described in the past a solid-phase extraction capillary electrophoresis (SPE-CE) system which was used with triple quadrupole and ion trap ESI-MS/MS systems for the identification of proteins at the low femtomole to subfemtomole sensitivity level [43, 44]. While this system is highly sensitive, its operation is labor-intensive and its operation has not been automated. In order to devise an analytical system with both the sensitivity of a CE and the level of automation of LC, we have constructed

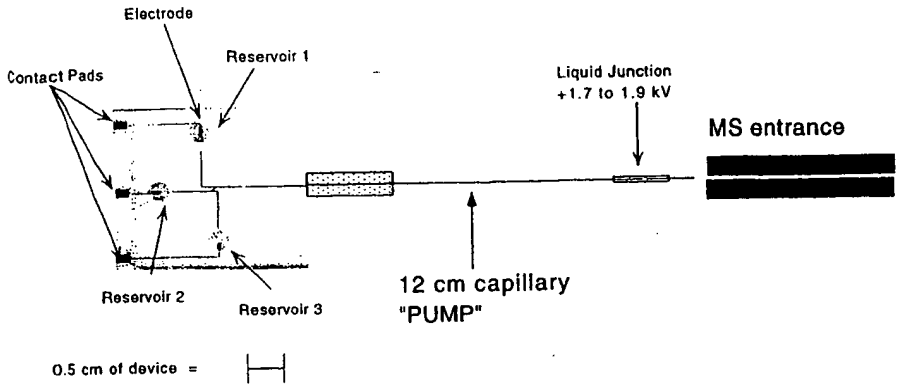


Figure 3. Schematic illustration of a microfabricated analytical system for CE, consisting of a micromachined device, coated capillary electroosmotic pump, and microelectrospray interface. The dimensions of the channels and reservoir are as indicated in the text. The channels on the device were graphically enhanced to make them more visible. Reproduced from [45], with permission.

microfabricated devices for the introduction of samples into ESI-MS for high-sensitivity peptide analysis.

The basic device is a piece of glass into which channels of 10–30 μm in depth and 50–70 μm in diameter are etched by using photolithography/etching techniques similar to the ones used in the semiconductor industry. (A simple device is shown in Fig. 3). The channels are connected to an external high voltage power supply [45]. Samples are manipulated on the device and off the device to the MS by applying different potentials to the reservoirs. This creates a solvent flow by electroosmotic pumping which can be redirected by changing the position of the electrode. Therefore, without the need for valves or gates and without any external pumping, the flow can be redirected by simply switching the position of the electrodes on the device. The direction and rate of the flow can be modulated by the size and the polarity of the electric field applied and also by the charge state of the surface.

The type of data generated by the system is illustrated in Fig. 4, which shows the mass spectrum of a peptide sample representing the tryptic digest of carbonic anhydrase at 290 fmol/μL. Each numbered peak indicates a peptide successfully identified as being derived from carbonic anhydrase. Some of the unassigned signals may be chemical or peptide contaminants. The MS is programmed to automatically select each peak and subject the peptide to CID. The resulting CID spectra are then used to identify the protein by correlation with sequence databases. Therefore, this system allows us to concurrently apply a number of protein digests onto the device, to sequentially mobilize the samples, to automatically generate CID spectra of selected peptide ions and to search sequence databases for protein identification. These steps are performed automatically without the need for user input and proteins can be identified at very low femtomole level sensitivity at a rate of approximately one protein per 15 min.

3.4 Assessment of 2-DE-MS proteome technology

Using a combination of the analytical techniques described above we have identified the 80 protein spots indicated in Fig. 5. The protein pattern was generated by separating a total of 40 microgram of protein contained in a total cell lysate of the yeast strain YPH499 by high resolution 2-DE and silver staining of the separated proteins. To estimate how far this type of proteome analysis can penetrate towards the identification of low abundance proteins, we have calculated the codon bias of the genes encoding the respective proteins. Codon bias is a

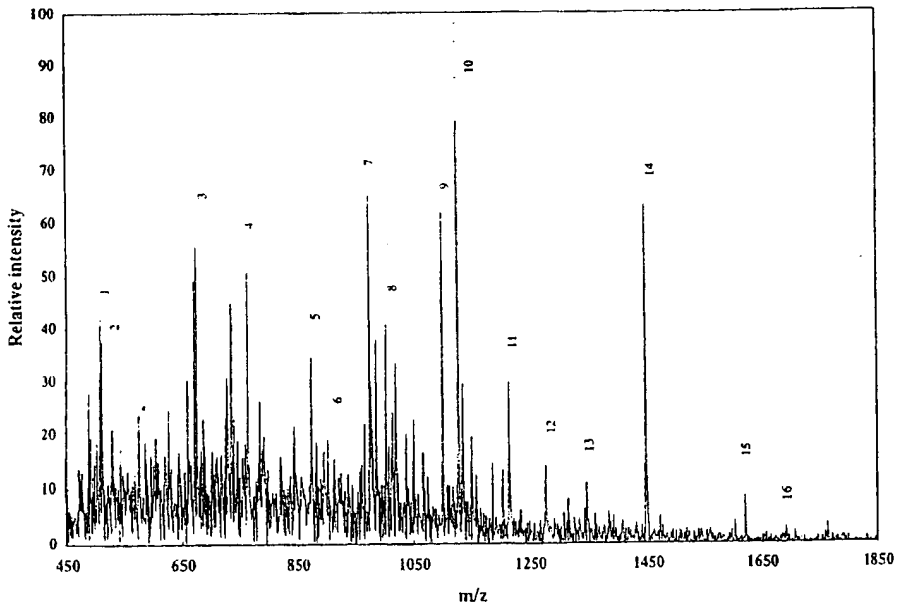


Figure 4. MS spectrum of a tryptic digest of carbonic anhydrase using the microfabricated system shown in Fig. 3. 290 fmol/μL of carbonic anhydrase tryptic digest was infused into a Finnigan LCQ ion trap MS. Each peak was selected for CID, and those which were identified as containing peptides derived from carbonic anhydrase are numbered. Reproduced from [45], with permission.

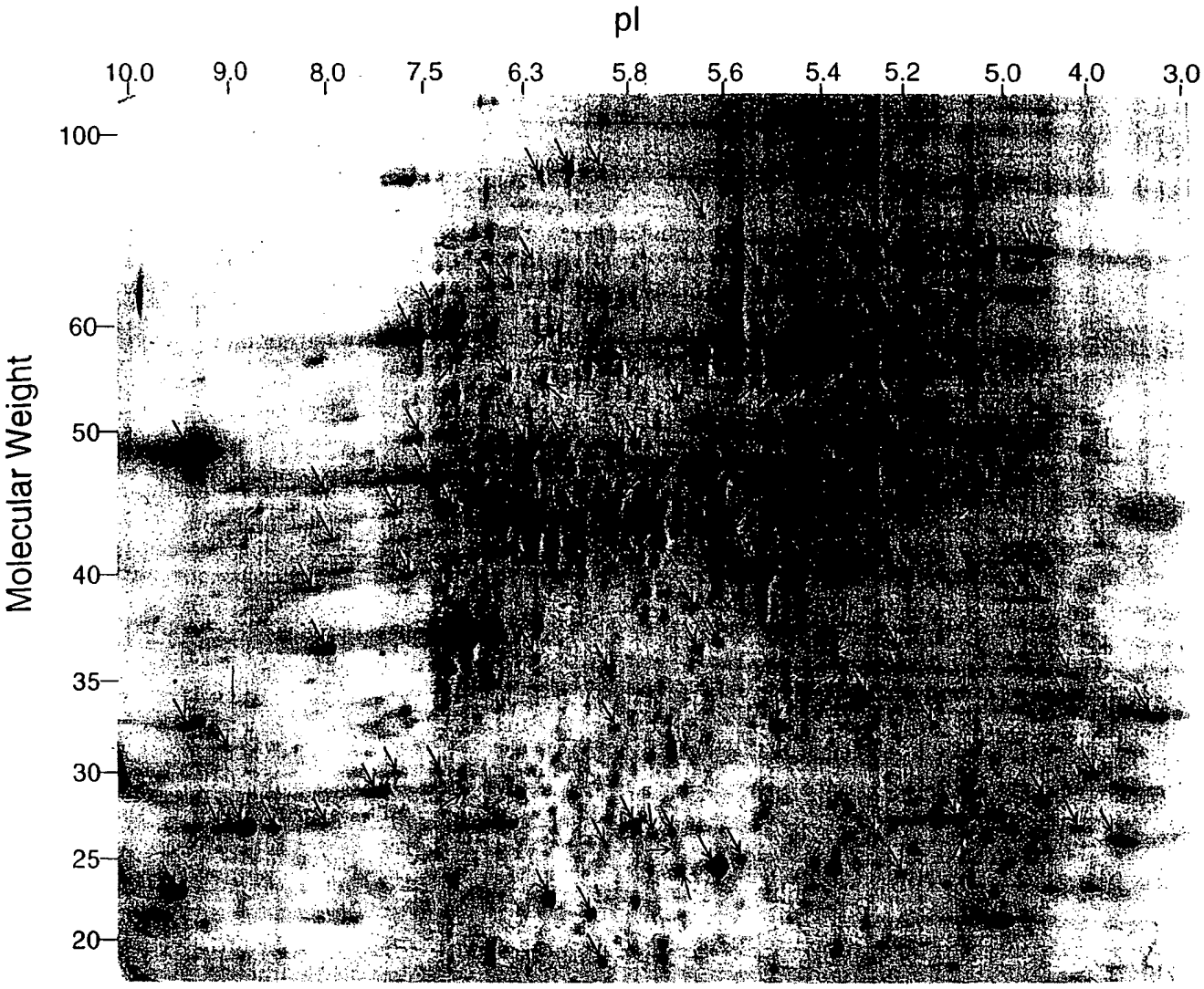


Figure 5. 2-DE separation of a lysate of yeast cells, with identified proteins highlighted. The first dimension of separation was an IPG from pH 3–10, and the second dimension was a 10%T SDS-PAGE gel. Proteins were visualized by silver staining. Further details of experimental procedures are included in S. P. Gygi *et al.* (submitted).

calculated measure of the degree of redundancy of triplet DNA codons used to produce each amino acid in a particular gene sequence. It has been shown to be a useful indicator of the level of the protein product of a particular gene sequence present in a cell [46]. The general rule which applies is that the higher the value of the codon bias calculated for a gene, the more abundant the protein product of that gene becomes. The calculated codon bias values corresponding to the proteins identified in Fig. 5 are shown in Fig. 6b. Nearly all of the proteins identified (> 95%) have codon bias values of > 0.2, indicating they are highly abundant in cells. In contrast, codon bias values calculated for the entire yeast genome (Fig. 6a) show that the majority of proteins present in the proteome have a codon bias of < 0.2 and are thus of low abundance.

This finding is of considerable importance in our assessment of the current status of proteome analysis technology. It is clear that even using highly sensitive analytical techniques, we are only able to visualize and identify the more abundant proteins. Since many important regulatory proteins are present only at low abundance, these would not be amenable to analysis using such techniques. This situation would be exacerbated in the analysis of proteomes containing many more proteins than the approximately 6000 gene products present in yeast cells [16]. In the analysis of, for example, the proteome of any human cells, there are potentially 50 000–100 000 gene products [47]. Inherent limitations on the amount of protein that can be loaded on 2-DE, and the number of components that can be resolved, indicate that only the most highly abundant fraction of the many gene products could be successfully analyzed. One approach that has been employed to circumvent these limitations is the use of very narrow range immobilized pH gradient strips for the first-dimension separation of 2-DE [48]. Since only those proteins which focus within the narrow range will enter the second dimension of separation, a much higher sample loading within the desired range is possible. This, in turn, can lead to the visualization and identification of less abundant proteins.

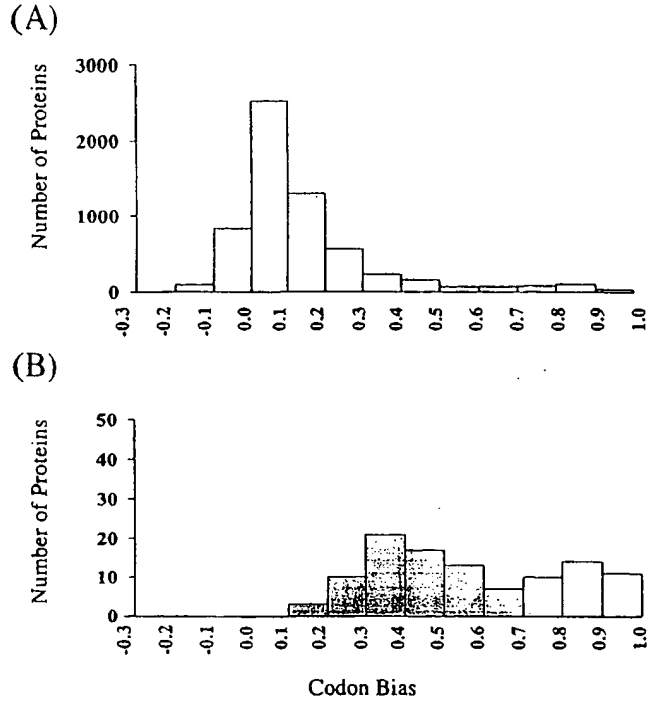


Figure 6. Calculated codon bias values for yeast proteins. (A) Distribution of calculated values for the entire yeast proteome. (B) Distribution of calculated values for the subset of 80 identified proteins also shown in Figs. 1 and 5. Further details of experimental procedures are included in S. P. Gygi *et al.* (submitted).

4 Utility of proteome analysis for biological research

For the success of proteomics as a mainstream approach to the analysis of biological systems it is essential to define how proteome analysis and biological research projects intersect. Without a clear plan for the implementation of proteome-type approaches into biological research projects the full impact of the technology can not be realized. The literature indicates that proteome analysis is used both as a database/data archive, and as a biological assay or biological research tool.

4.1 The proteome as a database

The use of proteomics as a database or data archive essentially entails an attempt to identify all the proteins in a cell or species and to annotate each protein with the known biological information that is relevant for each protein. The level of annotation can, of course, be extensive. The most common implementation of this idea is the separation of proteins by high resolution 2-DE, the identification of each detected protein spot and the annotation of the protein spots in a 2-DE gel database format. This approach is complicated by the fact that it is difficult to precisely define a proteome and to decide which proteome should be represented in the database. In contrast to the genome of a species, which is essentially static, the proteome is highly dynamic. Processes such as differentiation, cell activation and disease can all significantly change the proteome of a species. This is illustrated in Fig. 7. The figure shows two high-resolution 2-DE maps of proteins isolated from rat serum. Fig. 7A is from the serum of normal rats, while Fig. 7B is from the serum of rats in acute-phase serum after prior treatment with an inflammation-causing agent [49]. It is obvious that the protein patterns are significantly different in several areas, raising the question of exactly which proteome is being described.

Therefore, a comprehensive proteome database of a species or cell type needs to contain all of the parameters which describe the state and the type of the cells from which the proteins were extracted as well as the software tools to search the database with queries which reflect the dynamics of biological systems. A comprehensive proteome database should be capable of quantitatively describing the fate of each protein if specific systems and pathways are activated in the cell. Specifically, the quantity, the degree of modification, the subcellular location and the nature of molecules specifically interacting with a protein as well as the rate of change of these variables should be described. Using these admittedly stringent criteria, there is currently no comlete proteome database. A number of such databases are, however, in the process of being constructed. The most advanced among them, in our opinion, are the yeast protein database YPD [50] (accessible at http://www.ypd.com) and the human 2D-PAGE databases of the Danish Centre for Human Genome Research [12] (accessible at http://biobase.dk/cgi-bin/celis). While neither can be considered complete as not all of the potential gene products are identified, both contain extensive annotation of supplemental information for many of the spots which are positively identified in reference samples.

4.2 The proteome as a biological assay

The use of proteome analysis as a biological assay or research tool represents an alternative approach to integrating biology with proteomics. To investigate the state of a system, samples are subjected to a specific proceess that allows the quantitative or qualitative measurement of some of the variables which describe the system. In typical biochemical assays one variable (*e.g.*, enzyme activity) of a single component (*e.g.*, a particular enzyme) is measured. Using proteomics as an assay, multiple variables (*e.g.*, expression level, rate of synthesis, phosphorylation state, etc.) are measured concurrently on many (ideally all) of the proteins in a sample. The use of proteomics as an assay is a less far-reaching proposition than the construction of a comprehensive proteome database. It does, however, represent a pragmatic approach which can be adapted to investigate specific systems and pathways, as long as the interpretation of the results takes into account that with current technology not all of the variables which describe the system can be observed (see Section 3.4).

A common implementation of proteome analysis as a biological assay is when a 2-DE protein pattern generated from the analysis of an experimental sample is compared to an array of reference patterns representing different states of the system under investigation. The state of the experimental system at the time the sample was generated is therefore determined by the quantita-

tive comparative analysis of hundreds to a few thousand proteins. Comparative analysis of the 2-DE patterns furthermore highlights quantitative and qualitative differences in the protein profiles which correlate with the state of the system. For this type of analysis it is not essential that all the proteins are identified or even visualized, although the results become more informative as more proteins are compared. It is obvious, however, that the possibility to identify any protein deemed characteristic for a particular state dramatically enhances this approach by opening up new avenues for experimentation.

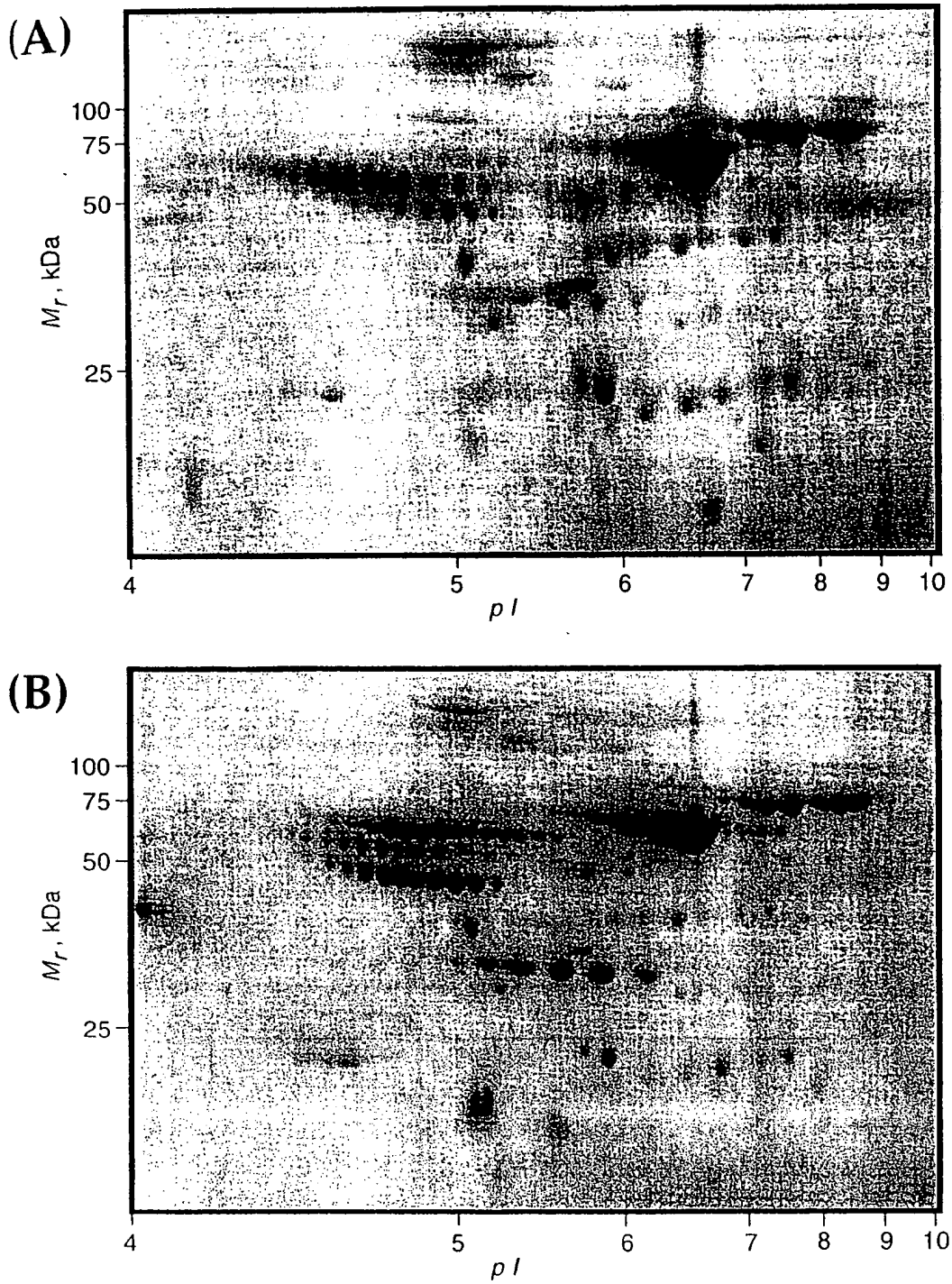


Figure 7. High resolution 2-DE map of proteins isolated from rat serum with or without prior exposure to an inflammation-causing agent. (A) normal rat serum, (B) acute-phase serum from rats which had previously been exposed to an inflammation-causing agent. The first dimension of separation is an IPG from pH 4–10, and the second dimension is a 7.5–17.5%T gradient SDS-PAGE gel. Proteins were visualized by staining with amido black. Further details of experimental procedures are included in [14, 49].

Proteome analysis as a biological assay has been successfully used in the field of toxicology, to characterize disease states or to study differential activation of cells. The approach is limited, of course, by the fact that only the visible protein spots are included in the assay, and it is well known that a substantial but far from complete fraction of cellular proteins are detected if a total cell lysate is separated by 2-DE. Proteins may not be detected in 2-DE gels because they are not abundant enough to be visualized by the detection method used, because they do not migrate within the boundaries (size, p*I*) resolved by the gel, because they are not soluble under the conditions used, or for other reasons.

A different way to use proteome analysis as a biological assay to define the state of a biological system is to take advantage of the wealth of information contained in 2-DE protein patterns. 2-DE is referred to as two-dimensional because of the electrophoretic mobility and the isoelectric points which define the position of each protein in a 2-DE pattern. In addition to the two dimensions used to generate the protein patterns, a number of additional data dimensions are contained in the protein patterns. Some of these dimensions such as protein expression level, phosphorylation state, subcellular location, association with other proteins, rate of synthesis or degradation indicate the activity state of a protein or a biological system. Comparative analysis of 2-DE protein patterns representing different states is therefore ideally suited for the detection, identification and analysis of suitable markers. Once again it must be emphasized that in this type of experiment only a fraction of the cellular proteins is analyzed. Since many regulatory proteins are of low abundance, this limitation is a concern, particularly in cases in which regulatory pathways are being investigated.

5 Concluding remarks

In this report we have addressed three main issues related to proteome analysis. First, we have discussed the rationale for studying proteomes. Second, we have assessed the technical feasibility of analyzing proteomes and described current proteome technology, and third, we have analyzed the utility of proteome analysis for biological research. It is apparent that proteome analysis is an essential tool in the analysis of biological systems. The multi-level control of protein synthesis and degradation in cells means that only the direct analysis of mature protein products can reveal their correct identities, their relevant state of modification and/or association and their amounts. Recently developed methods have enabled the identification of proteins at ever-increasing sensitivity levels and at a high level of automation of the analytical processes. A number of technical challenges, however, remain. While it is currently possible to identify essentially any protein spots that can be visualized by common staining methods, it is apparent that without prior enrichment only a relatively small and highly selected population of long-lived, highly expressed proteins is observed. There are many more proteins in a given cell which are not visualized by such methods. Frequently it is the low abundance proteins that execute key regulatory functions.

We have outlined the two principal ways proteome analysis is currently being used to intersect with biological research projects: the proteome as a database or data archive and proteome analysis as a biological assay. Both approaches have in common that at present they are conceptually and technically limited. Current proteome databases typically are limited to one cell type and one state of a cell and therefore do not account for the dynamics of biological systems. The use of proteome analysis as a biological assay can provide a wealth of information, but it is limited to the proteins detected and is therefore not truly proteome-wide. These limitations in proteomics are to a large extent a reflection of the fact that proteins in their fully processed form cannot easily be amplified and are therefore difficult to isolate in amounts sufficient for analysis or experimentation. The fact that to date no complete proteome has been described further attests to these difficulties. With continued rapid progress in protein analysis technology, however, we anticipate that the goal of complete proteome analysis will eventually become attainable.

We would like to acknowledge the funding for our work from the National Science Foundation Science and Technology Center for Molecular Biotechnology and from the NIH. We thank Yvan Rochon and Bob Franza for providing the yeast gel shown and Elisabetta Gianazza for providing the rat serum gels shown.

Received April 21, 1998

6 References

[1] Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J.-C., Yan, J. X., Gooley, A. A., Hughes, G., Humphery-Smith, I., Williams, K. L., Hochstrasser, D. F., *Bio/Technology* 1996, *14*, 61–65.

[2] Hodges, P. E., Payne, W. E., Garrels, J. I., *Nucleic Acids Res.* 1998, *26*, 68–72.

[3] O'Connor, C. D., Farris, M., Fowler, R., Qi, S. Y., *Electrophoresis* 1997, *18*, 1483–1490.

[4] Cordwell, S. J., Basseal, D. J., Humphery-Smith, I., *Electrophoresis* 1997, *18*, 1335–1346.

[5] Urquhart, B. L., Atsalos, T. E., Roach, D., Basseal, D. J., Bjellqvist, B., Britton, W. L., Humphery-Smith, I., *Electrophoresis* 1997, *18*, 1384–1392.

[6] Wasinger, V. C., Bjellqvist, B., Humphery-Smith, I., *Electrophoresis* 1997, *18*, 1373–1383.

[7] Link, A. J., Hays, L. G., Carmack, E. B., Yates III, J. R., *Electrophoresis* 1997, *18*, 1314–1334.

[8] Sazuka, T., Ohara, O., *Electrophoresis* 1997, *18*, 1252–1258.

[9] VanBogelen, R. A., Abshire, K. Z., Moldover, B., Olson, E. R., Neidhardt, F. C., *Electrophoresis* 1997, *18*, 1243–1251.

[10] Guerreiro, N., Redmond, J. W., Rolfe, B. G., Djordjevic, M. A., *Mol. Plant Microbe Interact.* 1997, *10*, 506–516.

[11] Yan, J. X., Tonella, L., Sanchez, J.-C., Wilkins, M. R., Packer, N. H., Gooley, A. A., Hochstrasser, D. F., Williams, K. L., *Electrophoresis* 1997, *18*, 491–497.

[12] Celis, J., Gromov, P., Ostergaard M., Madsen, P., Honoré, B., Dejgaard, K., Olsen, E., Vorum, H., Kristensen, D. B., Gromova, I., Haunso, A., Van Damme, J., Puype, M., Vandekerckhove, J., Rasmussen, H. H., *FEBS Lett.* 1996, *398*, 129–134.

[13] Appel, R. D., Sanchez, J.-C., Bairoch, A., Golaz, O., Miu, M., Vargas, J. R., Hochstrasser, D. F., *Electrophoresis* 1993, *14*, 1232–1238.

[14] Haynes, P., Miller, I., Aebersold, R., Gemeiner, M., Eberini, I., Lovati, R. M., Manzoni, C., Vignati, M., Gianazza, E., *Electrophoresis* 1998, *19*, 1484–1492.

[15] Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J.-F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J. D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, N. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, C. O., Venter, J. C., *Science* 1995, *269*, 496–512.

[16] Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S. G., *Science* 1996, *274*, 546.

[17] Fraser, C. M., Casjens, S., Huang, W. M., Sutton, G. G., Clayton, R., Lathigra, R., White, O., Ketchum, K. A., Dodson, R., Hickey, E. K., Gwinn, M., Dougherty, B., Tomb, J. F., Fleischmann, R. D., Richardson, D., Peterson, J., Kerlavage, A. R., Quackenbush, J., Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M. D., Gocayne, J., Weidman, J., Utterback, T., Watthey, T., McDonald, L., Artiach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M. D., Horst, K., Roberts, K., Hatch, B., Smith, H. O., Venter, J. C., *Nature* 1997, *390*, 580–586.

[18] Liang, P., Pardee, A. B., *Science* 1992, *257*, 967–971.

[19] Lashkari, D. A., DeRisi, J. L., McCusker, J. H., Namath, A. F., Gentile, C., Hwang, S. Y., Brown, P. O., Davis, R. W., *Proc. Natl. Acad. Sci. USA* 1997, *94*, 13057–13062.

[20] Shalon, D., Smith, S. J., Brown, P. O., *Genome Res.* 1996, *6*, 639–645.

[21] Velculescu, V. E., Zhang, L., Vogelstein, B., Kinzler, K. W., *Science* 1995, *270*, 484–487.

[22] Velculescu, V. E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M. A., Bassett, D. E., Hieter, P., Vogelstein, B., Kinzler, K. W., *Cell* 1997, *88*, 243–251.

[23] Krishna, R. G., Wold, F., *Adv. Enzymol.* 1993, *67*, 265–298.

[24] Görg, A., Postel, W., Gunther, S., *Electrophoresis* 1988, *9*, 531–546.

[25] Klose, J., Kobalz, U., *Electrophoresis* 1995, *16*, 1034–1059.

[26] Matsudaira, P., *J. Biol. Chem.* 1987, *262*, 10035–10038.

[27] Aebersold, R. H., Teplow, D. B., Hood, L. E., Kent, S. B., *J. Biol. Chem.* 1986, *261*, 4229–4238.

[28] Rosenfeld, J., Capdevielle, J., Guillemot, J. C., Ferrara, P., *Anal. Biochem.* 1992, *203*, 173–179.

[29] Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E., Kent, S. B., *Proc. Natl. Acad. Sci. USA* 1987, *84*, 6970–6974.

[30] Honoré, B., Leffers, H., Madsen, P., Celis, J. E., *Eur. J. Biochem.* 1993, *218*, 421–430.

[31] Mann, M., Wilm, M., *Anal. Chem.* 1994, *66*, 4390–4399.

[32] Eng, J., McCormack, A. L., Yates III, J. R., *J. Amer. Mass Spectrom.* 1994, *5*, 976–989.

[33] Yates III, J. R., Eng, J. K., McCormack, A. L., Schieltz, D., *Anal. Chem.* 1995, *67*, 1426–1436.

[34] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, *68*, 850–858.

[35] Hess, D., Covey, T. C., Winz, R., Brownsey, R. W., Aebersold, R., *Protein Sci.* 1993, *2*, 1342–1351.

[36] van Oostveen, I., Ducret, A., Aebersold, R., *Anal. Biochem.* 1997, *247*, 310–318.

[37] Lui, M., Tempst, P., Erdjument-Bromage, H., *Anal. Biochem.* 1996, *241*, 156–166.

[38] Patterson, S. D., Aebersold, R. A., *Electrophoresis* 1995, *16*, 1791–1814.

[39] Ducret, A., Foyn, Brunn, C., Bures, E. J., Marhaug, G., Husby, G. R. A., *Electrophoresis* 1996, *17*, 866–876.

[40] Haynes, P. A., Fripp, N., Aebersold, R., *Electrophoresis* 1998, *19*, 939–945.

[41] Figeys, D., Van Oostveen, I., Ducret, A., Aebersold, R., *Anal. Chem.* 1996, *68*, 1822–1828.

[42] Ducret, A., Van Oostveen, I., Eng, J. K., Yates III, J. R., Aebersold, R., *Protein Sci.* 1997, *7*, 706–719.

[43] Figeys, D., Ducret, A., Yates III, J. R., Aebersold, R., *Nature Biotech.* 1996, *14*, 1579–1583.

[44] Figeys, D., Aebersold, R., *Electrophoresis* 1997, *18*, 360–368.

[45] Figeys, D., Ning, Y., Aebersold, R., *Anal. Chem.* 1997, *69*, 3153–3160.

[46] Garrels, J. I., McLaughlin, C. S., Warner, J. R., Futcher, B., Latter, G. I., Kobayashi, R., Schwender, B., Volpe, T., Anderson, D. S., Mesquita-Fuentes, R., Payne, W. E., *Electrophoresis* 1997, *18*, 1347–1360.

[47] Schuler, G. D., Boguski, M. S., Stewart, E. A., Stein, L. D., Gyapay, G., Rice, K., White, R. E., Rodriguez-Tome, P., Aggarwal, A., Bajorek, E., Bentolila, S., Birren, B. B., Butler, A., Castle, A. B., Chiannilkulchai, N., Chu, A., Clee, C., Cowles, S., Day, P. J., Dibling, T., Drouot, N., Dunham, I., Duprat, S., Edwards, C., Fan, J.-B., Fang, N., Fizames, C., Garrett, C., Green, L., Hadley, D., Harris, M., Harrison, P., Brady, S., Hicks, A., Holloway, E., Hui, L., Hussain, S., Louis-Dit-Sully, C., Ma, J., MacGilvery, A., Mader, C., Maratukulam, A., Matise, T. C., McKusick, K. B., Morissette, J., Mungall, A., Muselet, D., Nusbaum, H. C., Page, D. C., Peck, A., Perkins, S., Piercy, M., Qin, F., Quackenbush, J., Ranby, S., Reif, T., Rozen, S., Sanders, X., She, X., Silva, J., Slonim, D. K., Soderlund, C., Sun, W.-L., Tabar, P., Thangarajah, T., Vega-Czarny, N., Vollrath, D., Voyticky, S., Wilmer, T., Wu, X., Adams, M. D., Auffray, C., Walter, N. A. R., Brandon, R., Dehejia, A., Goodfellow, P. N., Houlgatte, R., Hudson, J. R., Jr., Ide, S. E., Iorio, K. R., Lee, W. Y., Seki, N., Nagase, T., Ishikawa, K., Nomura, N., Phillips, C., Polymeropoulos, M. H., Sandusky, M., Schmitt, K., Berry, R., Swanson, K., Torres, R., Venter, J. C., Sikela, J. M., Beckmann, J. S., Weissenbach, J., Myers, R. M., Cox, D. R., James, M. R., Bentley, D., *et al. Science* 1996, *274*, 540–546.

[48] Sanchez, J.-C., Rouge, V., Pisteur, M., Ravier, F., Tonella, L., Moosmayer, M., Wilkins, M. R., Hochstrasser, D. F., *Electrophoresis* 1997, *18*, 324–327.

[49] Miller, I., Haynes, P., Gemeiner, M., Aebersold, R., Manzoni, C., Lovati, M. R., Vignati, M., Eberini, I., Gianazza, E., *Electrophoresis* 1998, *19*, 1493–1500.

[50] Garrels, J. I., *Nucleic Acids Res.* 1996, *24*, 46–49.

MOLECULAR AND CELLULAR BIOLOGY, Mar. 1999, p. 1720–1730 Vol. 19, No. 3
0270-7306/99/$04.00+0

Correlation between Protein and mRNA Abundance in Yeast

STEVEN P. GYGI, YVAN ROCHON, B. ROBERT FRANZA, AND RUEDI AEBERSOLD*

Department of Molecular Biotechnology, University of Washington, Seattle, Washington 98195-7730

Received 5 October 1998/Returned for modification 11 November 1998/Accepted 2 December 1998

We have determined the relationship between mRNA and protein expression levels for selected genes expressed in the yeast *Saccharomyces cerevisiae* growing at mid-log phase. The proteins contained in total yeast cell lysate were separated by high-resolution two-dimensional (2D) gel electrophoresis. Over 150 protein spots were excised and identified by capillary liquid chromatography-tandem mass spectrometry (LC-MS/MS). Protein spots were quantified by metabolic labeling and scintillation counting. Corresponding mRNA levels were calculated from serial analysis of gene expression (SAGE) frequency tables (V. E. Velculescu, L. Zhang, W. Zhou, J. Vogelstein, M. A. Basrai, D. E. Bassett, Jr., P. Hieter, B. Vogelstein, and K. W. Kinzler, Cell 88:243–251, 1997). We found that the correlation between mRNA and protein levels was insufficient to predict protein expression levels from quantitative mRNA data. Indeed, for some genes, while the mRNA levels were of the same value the protein levels varied by more than 20-fold. Conversely, invariant steady-state levels of certain proteins were observed with respective mRNA transcript levels that varied by as much as 30-fold. Another interesting observation is that codon bias is not a predictor of either protein or mRNA levels. Our results clearly delineate the technical boundaries of current approaches for quantitative analysis of protein expression and reveal that simple deduction from mRNA transcript analysis is insufficient.

The description of the state of a biological system by the quantitative measurement of the system constituents is an essential but largely unexplored area of biology. With recent technical advances including the development of differential display-PCR (21), of cDNA microarray and DNA chip technology (20, 27), and of serial analysis of gene expression (SAGE) (34, 35), it is now feasible to establish global and quantitative mRNA expression profiles of cells and tissues in species for which the sequence of all the genes is known. However, there is emerging evidence which suggests that mRNA expression patterns are necessary but are by themselves insufficient for the quantitative description of biological systems. This evidence includes discoveries of posttranscriptional mechanisms controlling the protein translation rate (15), the half-lives of specific proteins or mRNAs (33), and the intracellular location and molecular association of the protein products of expressed genes (32).

Proteome analysis, defined as the analysis of the protein complement expressed by a genome (26), has been suggested as an approach to the quantitative description of the state of a biological system by the quantitative analysis of protein expression profiles (36). Proteome analysis is conceptually attractive because of its potential to determine properties of biological systems that are not apparent by DNA or mRNA sequence analysis alone. Such properties include the quantity of protein expression, the subcellular location, the state of modification, and the association with ligands, as well as the rate of change with time of such properties. In contrast to the genomes of a number of microorganisms (for a review, see reference 11) and the transcriptome of *Saccharomyces cerevisiae* (35), which have been entirely determined, no proteome map has been completed to date.

The most common implementation of proteome analysis is the combination of two-dimensional gel electrophoresis (2DE) (isoelectric focusing-sodium dodecyl sulfate [SDS]-polyacrylamide gel electrophoresis) for the separation and quantitation of proteins with analytical methods for their identification. 2DE permits the separation, visualization, and quantitation of thousands of proteins reproducibly on a single gel (18, 24). By itself, 2DE is strictly a descriptive technique. The combination of 2DE with protein analytical techniques has added the possibility of establishing the identities of separated proteins (1, 2) and thus, in combination with quantitative mRNA analysis, of correlating quantitative protein and mRNA expression measurements of selected genes.

The recent introduction of mass spectrometric protein analysis techniques has dramatically enhanced the throughput and sensitivity of protein identification to a level which now permits the large-scale analysis of proteins separated by 2DE. The techniques have reached a level of sensitivity that permits the identification of essentially any protein that is detectable in the gels by conventional protein staining (9, 29). Current protein analytical technology is based on the mass spectrometric generation of peptide fragment patterns that are idiotypic for the sequence of a protein. Protein identity is established by correlating such fragment patterns with sequence databases (10, 22, 37). Sophisticated computer software (8) has automated the entire process such that proteins are routinely identified with no human interpretation of peptide fragment patterns.

In this study, we have analyzed the mRNA and protein levels of a group of genes expressed in exponentially growing cells of the yeast *S. cerevisiae*. Protein expression levels were quantified by metabolic labeling of the yeast proteins to a steady state, followed by 2DE and liquid scintillation counting of the selected, separated protein species. Separated proteins were identified by in-gel tryptic digestion of spots with subsequent analysis by microspray liquid chromatography-tandem mass spectrometry (LC-MS/MS) and sequence database searching. The corresponding mRNA transcript levels were calculated from SAGE frequency tables (35).

This study, for the first time, explores a quantitative comparison of mRNA transcript and protein expression levels for a relatively large number of genes expressed in the same metabolic state. The resultant correlation is insufficient for predic-

* Corresponding author. Mailing address: Department of Molecular Biotechnology, Box 357730, University of Washington, Seattle, WA 98195-7730. Phone: (206) 221-4196. Fax: (206) 685-7301. E-mail: ruedi@u.washington.edu.

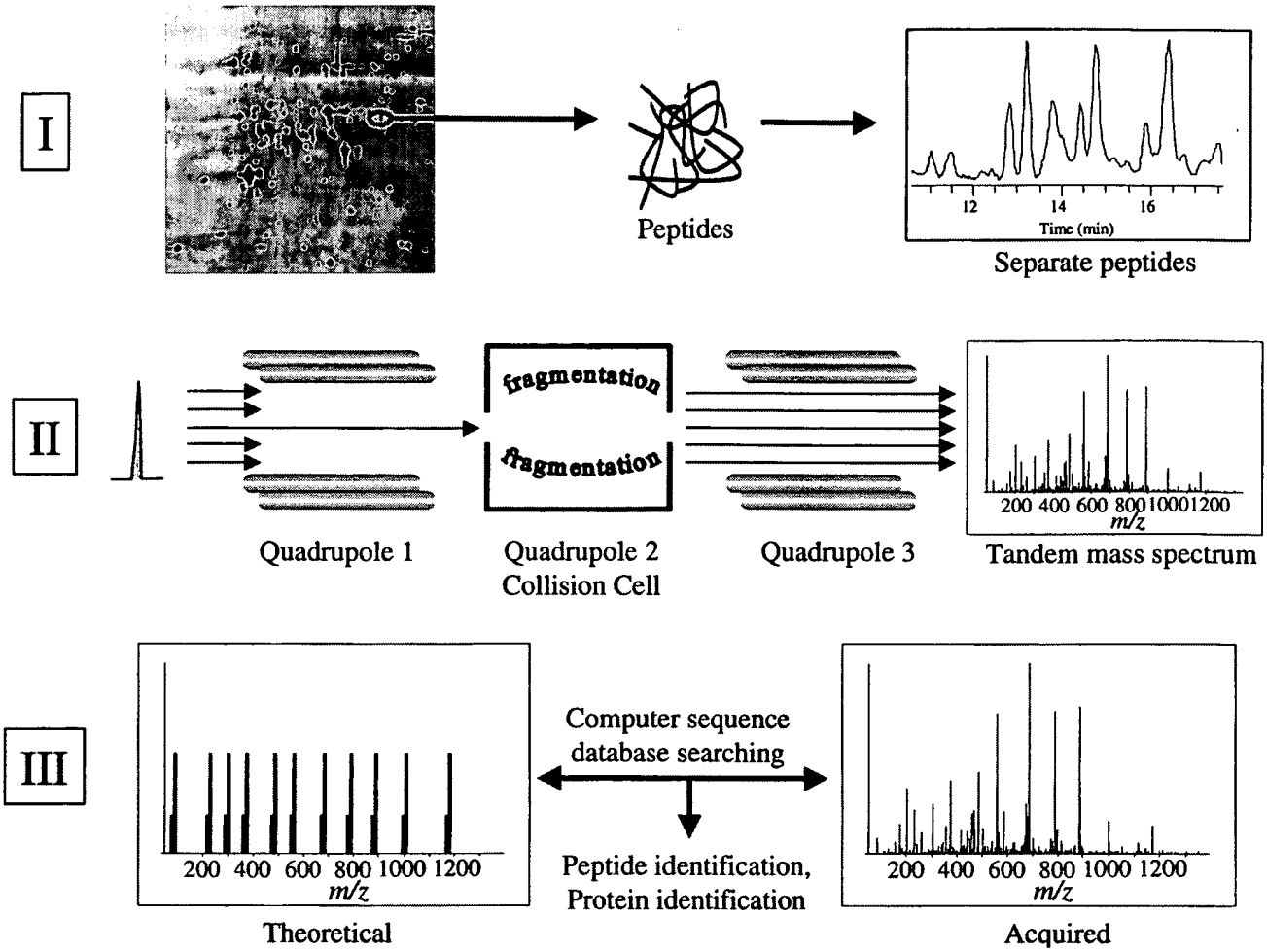


FIG. 1. Schematic illustration of proteome analysis by 2DE and mass spectrometry. In part I, proteins are separated by 2DE, stained spots are excised and subjected to in-gel digestion with trypsin, and the resulting peptides are separated by on-line capillary high-performance liquid chromatography. In part II, a peptide is shown eluting from the column in part I. The peptide is ionized by electrospray ionization and enters the mass spectrometer. The mass of the ionized peptide is detected, and the first quadrupole mass filter allows only the specific mass-to-charge ratio of the selected peptide ion to pass into the collision cell. In the collision cell, the energized, ionized peptides collide with neutral argon gas molecules. Fragmentation of the peptide is essentially random but occurs mainly at the peptide bonds, resulting in smaller peptides of differing lengths (masses). These peptide fragments are detected as a tandem mass (MS/MS) spectrum in the third quadrupole mass filter where two ion series are recorded simultaneously, one each from sequencing inward from the N and C termini of the peptide, respectively. In part III, the MS/MS spectrum from the selected, ionized peptide is compared to predicted tandem mass spectra computer generated from a sequence database. Provided that the peptide sequence exists in the database, the peptide and, by association, the protein from which the peptide was derived can be identified. Unambiguous protein identification is attained in a single analysis because multiple peptides are identified as being derived from the same protein.

tion of protein levels from mRNA transcript levels. We have also compared the relative amounts of protein and mRNA with the respective codon bias values for the corresponding genes. This comparison indicates that codon bias by itself is insufficient to accurately predict either the mRNA or the protein expression levels of a gene. In addition, the results demonstrate that only highly expressed proteins are detectable by 2DE separation of total cell lysates and that therefore the construction of complete proteome maps with current technology will be very challenging, irrespective of the type of organism.

MATERIALS AND METHODS

Yeast strain and growth conditions. The source of protein and message transcripts for all experiments was YPH499 (*MATa ura3-52 lys2-801 ade2-101 leu2-Δ1 his3-Δ200 trp1-Δ63*) (30). Logarithmically growing cells were obtained by growing yeast cells to early log phase (3×10^6 cells/ml) in YPD rich medium (YPD supplemented with 6 mM uracil, 4.8 mM adenine, and 24 mM tryptophan) at 30°C (35). Metabolic labeling of protein was accomplished in YPD medium exactly as described elsewhere (4) with the exception that 1 ml of cells was labeled with 3 mCi to offset methionine present in YPD medium. Protein was harvested as described by Garrels and coworkers (12). Harvested protein was lyophilized, resuspended in isoelectric focusing gel rehydration solution, and stored at −80°C.

2DE. Soluble proteins were run in the first dimension by using a commercial flatbed electrophoresis system (Multiphor II; Pharmacia Biotech). Immobilized polyacrylamide gel (IPG) dry strips with nonlinear pH 3.0 to 10.0 gradients (Amersham-Pharmacia Biotech) were used for the first-dimension separation. Forty micrograms of protein from whole-cell lysates was mixed with IPG strip rehydration buffer (8 M urea, 2% Nonidet P-40, 10 mM dithiothreitol), and 250 to 380 μl of solution was added to individual lanes of an IPG strip rehydration tray (Amersham-Pharmacia Biotech). The strips were allowed to rehydrate at room temperature for 1 h. The samples were run at 300 V–10 mA–5 W for 2 h, then ramped to 3,500 V–10 mA–5 W over a period of 3 h, and then kept at 3,500 V–10 mA–5 W for 15 to 19 h. At the end of the first-dimension run (60 to 70 kV · h), the IPG strips were reequilibrated for 8 min in 2% (wt/vol) dithiothreitol in 2% (wt/vol) SDS–6 M urea–30% (wt/vol) glycerol–0.05 M Tris HCl (pH 6.8) and for 4 min in 2.5% iodoacetamide in 2% (wt/vol) SDS–6 M urea–30% (wt/vol) glycerol–0.05 M Tris HCl (pH 6.8). Following reequilibration, the strips were transferred and apposed to 10% polyacrylamide second-dimension gels. Polyacrylamide gels were poured in a casting stand with 10% acrylamide–2.67% piperazine diacrylamide–0.375 M Tris base-HCl (pH 8.8)–0.1% (wt/vol) SDS–0.05%

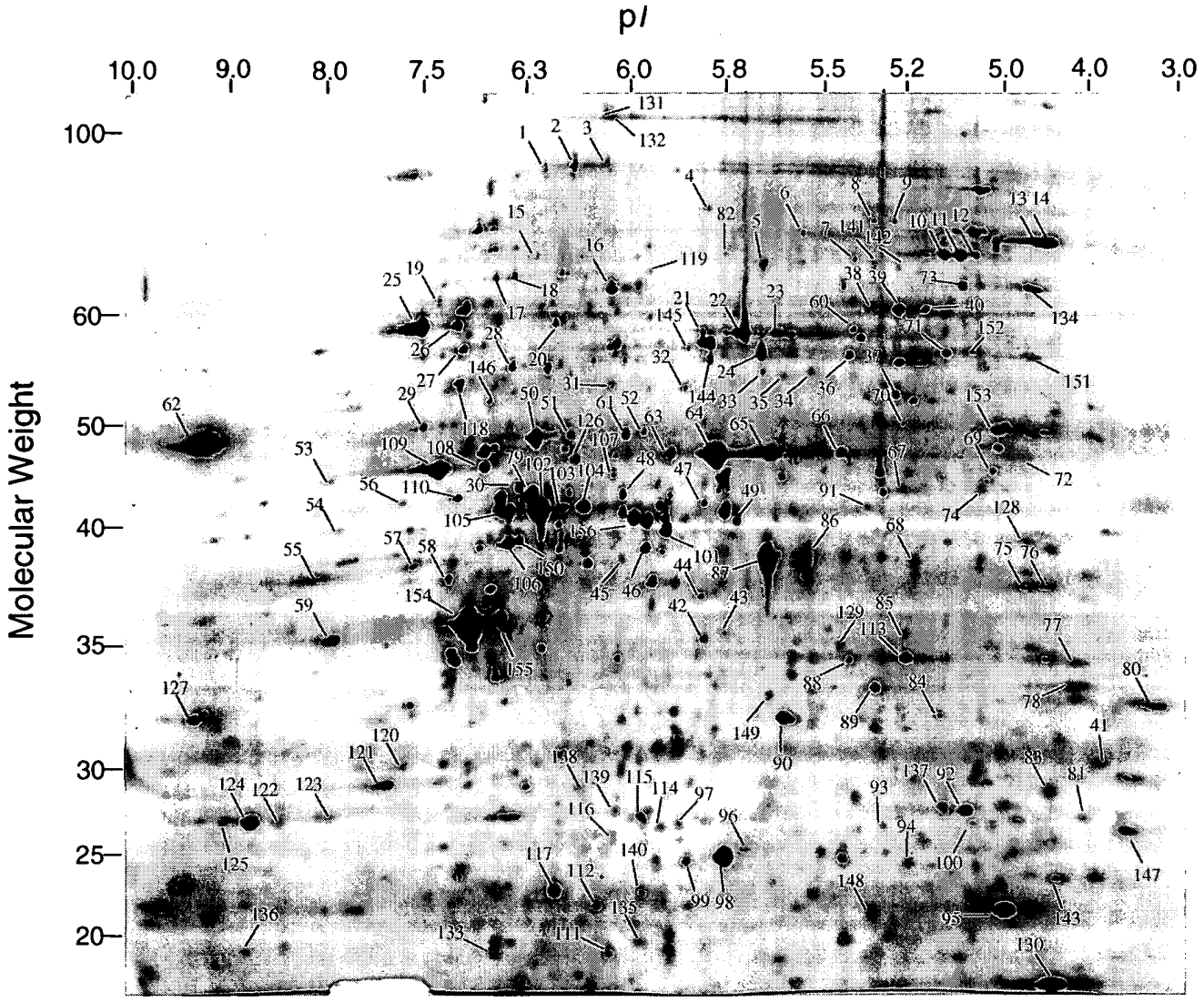


FIG. 2. 2D silver-stained gel of the proteins in yeast total cell lysate. Proteins were separated in the first dimension (horizontal) by isoelectric focusing and then in the second dimension (vertical) by molecular weight sieving. Protein spots (156) were chosen to include the entire range of molecular weights, isoelectric focusing points, and staining intensities. Spots were excised, and the corresponding protein was identified by mass spectrometry and database searching. The spots are labeled on the gel and correspond to the data presented in Table 1. Molecular weights are given in thousands.

(wt/vol) ammonium persulfate–0.05% TEMED (*N,N,N',N'*-tetramethylethylenediamine) in Milli-Q water. The apparatus used to run second-dimension gels was a noncommercial apparatus from Oxford Glycosciences, Inc. Once the IPG strips were apposed to the second-dimension gels, they were immediately run at 50 mA (constant)–500 V–85 W for 20 min, followed by 200 mA (constant)–500 V–85 W until the buffer front line was 10 to 15 mm from the bottom of the gel. Gels were removed and silver stained according to the procedure of Shevchenko et al. (29).

Protein identification. Gels were exposed to X-ray film overnight, and then the silver staining and film were used to excise 156 spots of varying intensities, molecular weights, and isoelectric focusing points. In order to increase the detection limit by mass spectrometry, spots were cut out and pooled from up to four identical cold, silver-stained gels. In-gel tryptic digests of pooled spots were performed as described previously (29). Tryptic peptides were analyzed by microcapillary LC-MS with automated switching to MS/MS mode for peptide fragmentation. Spectra were searched against the composite OWL protein sequence database (version 30.2; 250,514 protein sequences) (24a) by using the computer program Sequest (8), which matches theoretical and acquired tandem mass spectra. A protein match was determined by comparing the number of peptides identified and their respective cross-correlation scores. All protein identifications were verified by comparison with theoretical molecular weights and isoelectric points.

mRNA quantitation. Velculescu and coworkers have previously generated frequency tables for yeast mRNA transcripts from the same strain grown under the same stated conditions as described herein (35). The SAGE technology is based on two main principles. First, a short sequence tag (15 bp) that contains sufficient information uniquely to identify a transcript is generated. A single tag is usually generated from each mRNA transcript in the cell which corresponds to 15 bp at the 3'-most cutting site for *Nla*III. Second, many transcript tags can be concatenated into a single molecule and then sequenced, revealing the identity of multiple tags simultaneously. Over 20,000 transcripts were sequenced from yeast strain YPH499 growing at mid-log phase on glucose. Assuming the previously derived estimate of 15,000 mRNA molecules per cell (16), this would represent a 1.3-fold coverage even for mRNA molecules present at a single copy per cell and would provide a 72% probability of detecting such transcripts. Computer software which took for input the gene detected, examined the nucleotide sequence, and performed the calculation as described by Velculescu and coworkers (35) was written. In practice, we found that for 21 of 128 (16%) genes examined viable mRNA levels from SAGE data could not be calculated. This was because (i) no CATG site was found in the open reading frame (ORF), (ii) a CATG site was found but the corresponding 10-bp putative SAGE tag was not found in the frequency tables, or (iii) identical putative SAGE tags were present for multiple genes (e.g., TDH2_YEAST and TDH3_YEAST).

TABLE 1. Expressed genes identified from 2D gel in Fig. 2

Mol wt	pI	Spot no.	YPD gene name[a]	Protein abundance (10^3 copies/cell)	mRNA abundance (copies/cell)	Codon bias
17,259	6.75	133	CPR1	15.2	61.7	0.769
18,702	4.80	83	EGD2	20.1	5.2	0.724
18,726	4.44	147	YKL056C	61.2	88.4	0.831
18,978	5.95	135	YER067W	3.7	6.7	0.118
19,108	5.04	130	YLR109W	94.4	9.7	0.680
19,681	9.08	136	ATP7	11.0	NA[b,c]	0.246
20,505	6.07	111	GUK1	16.5	3.7	0.422
21,444	5.25	148	SAR1	5.4	10.4	0.455
21,583	4.98	95	TSA1	110.6	40.1	0.845
22,602	4.30	80	EFB1	66.1	23.8	0.875
23,079	6.29	112	SOD2	12.6	2.2	0.351
23,743	5.44	137	HSP26	NA[d]	0.7	0.434
24,033	5.97	96	ADK1	17.4	16.4	0.656
24,058	4.43	143	YKL117W	29.2	10.4	0.339
24,353	6.30	140	TFS1	8.1	0.7	0.146
24,662	5.85	99	URA5	25.4	6.0	0.359
24,808	6.33	97	GSP1	26.3	5.2	0.735
24,908	8.73	122	RPS5	18.6	NA[c]	0.899
25,081	4.65	81	MRP8	9.3	NA[c]	0.241
25,960	6.06	116	RPE1	5.8	0.7	0.372
26,378	9.55	127	RPS3	96.8	NA[c]	0.863
26,467	5.18	100	VMA4	10.5	3.7	0.427
26,661	5.84	98	TPI1	NA[d]	NA[c]	0.900
27,156	5.56	93	PRE8	6.9	0.7	0.129
27,334	6.13	115	YHR049W	18.4	2.2	0.520
27,472	5.33	92	YNL010W	31.6	3.7	0.421
27,480	8.95	123	GPM1	10.0	169.4	0.902
27,480	8.95	124	GPM1	231.4	169.4	0.902
27,480	8.95	125	GPM1	7.5	169.4	0.902
27,809	5.97	139	HOR2	5.7	0.7	0.381
27,874	4.46	78	YST1	13.6	52.8	0.805
28,595	4.51	41	PUP2	4.4	0.7	0.147
29,156	6.59	114	YMR226C	14.5	2.2	0.283
29,244	8.40	120	DPM1	5.0	11.2	0.362
29,443	5.91	48	PRE4	3.4	3.7	0.162
30,012	6.39	138	PRB1	21.2	1.5	0.449
30,073	4.63	77	BMH1	14.7	28.2	0.454
30,296	7.94	121	OMP2	67.4	41.6	0.499
30,435	6.34	89	GPP1	70.2	11.2	0.703
31,332	5.57	88	ILV6	13.9	3.0	0.402
32,159	5.46	113	IPP1	63.1	3.7	0.752
32,263	6.00	149	HIS1	22.4	4.5	0.232
33,311	5.35	84	SPE3	15.1	6.7	0.468
34,465	5.60	129	ADE1	8.7	5.2	0.305
34,762	5.32	85	SEC14	10.9	6.0	0.373
34,797	5.85	42	URA1	49.5	8.9	0.237
34,799	6.04	90	BEL1	103.2	81.0	0.875
35,556	5.97	43	YDL124W	6.4	4.5	0.206
35,619	8.41	59	TDH1	69.8	32.7[e]	0.940
35,650	5.49	68	CAR1	5.2	3.0	0.339
35,712	6.72	117	TDH2	49.6	473.0[e]	0.982
35,712	6.72	154	TDH2	863.5	473.0[e]	0.982
35,712	6.72	155	TDH2	79.4	473.0[e]	0.982
36,272	4.85	128	APA1	8.7	0.7	0.425
36,358	5.05	75	YJR105W	17.6	17.1	0.522
36,358	5.05	76	YJR105W	27.5	17.1	0.522
36,596	6.37	79	ADH2	58.9	260.0[e]	0.711
36,714	6.30	102	ADH1	746.1	260.0	0.913
36,714	6.30	103	ADH1	17.6	260.0	0.913
36,714	6.30	104	ADH1	61.4	260.0	0.913
36,714	6.30	105	ADH1	52.7	260.0	0.913
37,033	6.23	44	TAL1	44.8	3.7	0.701
37,796	7.36	57	IDH2	29.4	6.7	0.330
37,886	6.49	106	ILV5	76.0	4.5	0.892
38,700	7.83	55	BAT1	30.9	11.2	0.469
38,702	6.24	46	QCR2	NA[d]	2.2	0.326
39,477	5.58	86	FBA1	17.8	183.6	0.935
39,477	5.58	87	FBA1	427.2	183.6	0.935
39,540	6.50	150	HOM2	60.3	4.5	0.592
39,561	6.12	156	PSA1	96.4	27.5	0.718
41,158	6.01	49	YNL134C	14.9	1.5	0.316
41,623	7.18	58	BAT2	19.0	8.9	0.250
41,728	7.29	110	ERG10	24.1	4.5	0.543
41,900	5.42	74	TOM40	22.3	2.2	0.375
42,402	6.29	45	CYS3	6.7	8.9	0.621
42,883	5.63	67	DYS1	15.8	5.2	0.526
43,409	6.31	107	SER1	10.5	1.5	0.292
43,421	5.59	91	ERG6	2.2	14.1	0.408
44,174	7.32	56	YBR025C	13.1	6.0	0.684
44,682	4.99	72	TIF1	2.9	39.4	0.834
44,707	7.77	108	PGK1	23.7	165.7	0.897
44,707	7.77	109	PGK1	315.2	165.7	0.897
46,080	6.72	30	CAR2	15.4	NA[c]	0.495
46,383	8.52	53	IDP1	7.7	0.7	0.436
46,553	5.98	47	IDP2	32.4	NA[c]	0.197
46,679	6.39	50	ENO1	35.4	0.7	0.930
46,679	6.39	51	ENO1	6.6	0.7	0.930
46,679	6.39	52	ENO1	2.2	0.7	0.930
46,773	5.82	63	ENO2	15.5	289.1	0.960
46,773	5.82	64	ENO2	635.5	289.1	0.960
46,773	5.82	65	ENO2	93.0	289.1	0.960
46,773	5.82	66	ENO2	31.0	289.1	0.960
47,402	6.09	126	COR1	2.5	0.7	0.422
47,666	8.98	54	AAT2	11.7	6.0	0.338
48,364	5.25	73	WTM1	74.5	13.4	0.365
48,530	6.20	61	MET17	38.1	29.0	0.576
48,904	5.18	69	LYS9	16.2	3.7	0.463
48,987	4.90	153	SUP45	29.6	11.9	0.377
49,727	5.47	70	PRO2	13.6	5.2	0.297
49,912	9.27	62	TEF2	558.5	282.0	0.932
50,444	5.67	35	YDR190C	4.8	2.2	0.228
50,837	6.11	32	YEL047C	3.8	1.5	0.387
50,891	4.59	151	TUB2	11.2	7.4	0.404
51,547	6.80	27	LPD1	18.9	2.2	0.351
52,216	7.25	29	SHM2	19.7	7.4	0.722
52,859	5.54	37	YFR044C	30.2	6.7	0.442
53,798	5.19	71	HXK2	26.5	7.4	0.756
53,803	6.05	145	GYP6	4.4	0.7	0.147
54,403	5.29	39	ALD6	37.7	2.2	0.664
54,403	5.29	40	ALD6	6.6	2.2	0.664
54,502	6.20	31	ADE13	6.3	1.5	0.417
54,543	7.75	25	PYK1	225.3	101.8	0.965
54,543	7.75	26	PYK1	39.8	101.8	0.965
55,221	6.66	146	YEL071W	16.3	3.0	0.244
55,295	4.35	134	PDI1	66.2	14.1	0.589
55,364	5.98	24	GLK1	22.6	6.0	0.237
55,481	7.97	118	ATP1	21.6	2.2	0.637
55,886	6.47	28	CYS4	22.2	NA[c]	0.444
56,167	5.83	33	ARO8	14.3	3.0	0.324
56,167	5.83	34	ARO8	9.1	3.0	0.324
56,584	6.36	20	CYB2	18.9	NA[c]	0.259
57,366	5.53	60	FRS2	2.3	0.7	0.451
57,383	5.98	144	ZWF1	5.6	0.7	0.215
57,464	5.49	36	THR4	21.4	3.7	0.508
57,512	5.50	7	SRV2	6.5	NA[c]	0.260
57,727	4.92	152	VMA2	33.7	8.9	0.546
58,573	6.47	17	ACH1	4.4	1.5	0.327
58,573	6.47	18	ACH1	5.4	1.5	0.327
61,353	5.87	21	PDC1	6.5	200.7	0.962
61,353	5.87	22	PDC1	303.2	200.7	0.962
61,353	5.87	23	PDC1	16.3	200.7	0.962
61,649	5.54	38	CCT8	2.2	1.5	0.271

Continued

TABLE 1—*Continued*

Continued on following page

TABLE 1—*Continued*

Mol wt	pI	Spot no.	YPD gene name[a]	Protein abundance (10^3 copies/cell)	mRNA abundance (copies/cell)	Codon bias
61,902	6.21	101	PDC5	4.3	NA[c]	0.828
62,266	6.19	16	ICL1	20.1	NA[c]	0.327
62,862	8.02	19	ILV3	5.3	4.5	0.548
63,082	6.40	119	PGM2	2.2	3.0	0.402
64,335	5.77	5	PAB1	30.4	1.5	0.616
66,120	5.42	8	STI1	6.7	0.7	0.313
66,120	5.42	9	STI1	6.4	0.7	0.313
66,450	5.29	141	SSB2	7.0	NA[c]	0.880
66,450	5.29	142	SSB2	2.3	NA[c]	0.880
66,456	5.23	10	SSB1	64.5	79.5	0.907
66,456	5.23	11	SSB1	59.0	79.5	0.907
66,456	5.23	12	SSB1	13.7	79.5	0.907
68,397	5.82	82	LEU4	3.1	3.0	0.407
69,313	4.90	13	SSA2	24.3	18.6	0.892
69,313	4.90	14	SSA2	77.1	18.6	0.892
74,378	8.46	15	YKL029C	2.8	3.7	0.353
75,396	5.82	6	GRS1	5.5	7.4	0.500
85,720	6.25	1	MET6	2.0	NA[c]	0.772
85,720	6.25	2	MET6	10.9	NA[c]	0.772
85,720	6.25	3	MET6	1.4	NA[c]	0.772
93,276	6.11	131	EFT1	17.9	41.6	0.890
93,276	6.11	132	EFT1	5.7	41.6	0.890
102,064[e]	6.61[e]	94	ADE3	4.8	5.2	0.423
107,482[e]	5.33[e]	4	MCM3	2.7	NA[c]	0.240

[a] YPD gene names are available from the YPD website (39).
[b] NA, calculation could not be performed or was not available.
[c] mRNA data inconclusive or NA.
[d] No methionines in predicted ORF; therefore, protein concentration was not determined.
[e] Measured molecular weight or pI did not match theoretical molecular weight or pI.

Protein quantitation. [^{35}S]methionine-labeled gels were exposed to X-ray film overnight, and then the silver stain and film were used to excise 156 spots of varying intensities, molecular weights, and pIs. The excised spots were placed in 0.6-ml microcentrifuge tubes, and scintillation cocktail (100 μl) was added. The samples were vortexed and counted. In addition, two parallel gels were electroblotted to polyvinylidene difluoride membranes. The membranes were exposed to X-ray film, and four intense single spots were excised from each membrane and subjected to amino acid analysis. For these four spots, a mean of 209 ± 4 cpm/pmol of protein/methionine was found. This number was used to quantitate all remaining spots in conjunction with the number of methionines present in the protein.

To ensure that proteins were labeled to equilibrium, parallel 2D gels were prepared and run on yeast metabolically labeled for 1, 2, 6, or 18 h. The corresponding 156 spots were excised from each gel, and radioactivity was measured by liquid scintillation counting for each spot. Calculated protein levels were highly reproducible for all time points measured after 1 h.

Calculation of codon bias and predicted half-life. Codon bias values were extracted from the YPD spreadsheet (17). Protein half-lives were calculated based on the N-end rule (33). When the N-terminal processing was not known experimentally, it was predicted based on the affinity of methionine aminopeptidase (31).

RESULTS

Characteristics of proteome approach. Nearly every facet of proteome analysis hinges on the unambiguous identification of large numbers of expressed proteins in cells. Several techniques have been described previously for the identification of proteins separated by 2DE, including N-terminal and internal sequencing (1, 2), amino acid analysis (38), and more recently mass spectrometry (25). We utilized techniques based on mass spectrometry because they afford the highest levels of sensitivity and provide unambiguous identification. The specific procedure used is schematically illustrated in Fig. 1 and is based on three principles. First, proteins are removed from the gel by proteolytic in-gel digestion, and the resulting peptides are separated by on-line capillary high-performance liquid chromatography. Second, the eluting peptides are ionized and detected, and the specific peptide ions are selected and fragmented by the mass spectrometer. To achieve this, the mass spectrometer switches between the MS mode (for peptide mass identification) and the MS/MS mode (for peptide characterization and sequencing). Selected peptides are fragmented by a process called collision-induced dissociation (CID) to generate a tandem mass spectrum (MS/MS spectrum) that contains the peptide sequence information. Third, individual CID mass spectra are then compared by computer algorithms to predicted spectra from a sequence database. This results in the identification of the peptide and, by association, the protein(s) in the spot. Unambiguous protein identification is attained in a single analysis by the detection of multiple peptides derived from the same protein.

Protein identification. Yeast total cell protein lysate (40 μg), metabolically labeled with [^{35}S]methionine, was electrophoretically separated by isoelectric focusing in the first dimension and by SDS–10% polyacrylamide gel electrophoresis in the second dimension. Proteins were visualized by silver staining and by autoradiography. Of the more than 1,000 proteins visible by silver staining, 156 spots were excised from the gel and subjected to in-gel tryptic digestion, and the resulting peptides were analyzed and identified by microspray LC-MS/MS techniques as described above. The proteins in this study were all identified automatically by computer software with no human interpretation of mass spectra. They are indicated in Fig. 2 and detailed in Table 1.

The CID spectra shown in Fig. 3 indicate that the quality of the identification data generated was suitable for unambiguous protein identification. The spectra represent the amino acid sequences of tryptic peptides NSGDIVNLGSIAGR (Fig. 3A) and FAVGAFTDSLR (Fig. 3B). Both peptides were derived from protein S57593 (hypothetical protein YMR226C), which migrated to spot 114 (molecular weight, 29,156; pI, 6.59) in the 2D gel in Fig. 2. Five other peptides from the same analysis were also computer matched to the same protein sequence.

Protein and mRNA quantitation. For the 156 genes investigated, the protein expression levels ranged from 2,200 (PGM2) to 863,000 (TDH2/TDH3) copies/cell. The levels of mRNA for each of the genes identified were calculated from SAGE frequency tables (35). These tables contain the mRNA levels for 4,665 genes in yeast strain YPH499 grown to mid-log phase in YPD medium on glucose as a carbon source. In some instances, the mRNA levels could not be calculated for reasons stated in Materials and Methods. For the proteins analyzed in this study, mean transcript levels varied from 0.7 to 473 copies/cell.

Selection of the sample population for mRNA-protein expression level correlation. The protein spots selected for identification were selected from spots visible by silver staining in the 2D gel. An attempt was made not to include spots where overlap with other spots was readily apparent. The number of proteins identified was 156 (Table 1). Some proteins migrated to more than one spot (presumably due to differential protein processing or modifications), and protein levels from these spots were calculated by integrating the intensities of the different spots. The 156 protein spots analyzed represented the products of 128 different genes. Genes were excluded from the correlation analysis only if part of the data set was missing; i.e., genes were excluded if (i) no mRNA expression data were available for the protein or putative SAGE tags were ambiguous, (ii) the amino acid sequence did not contain methionine, (iii) more than a single protein was conclusively identified as

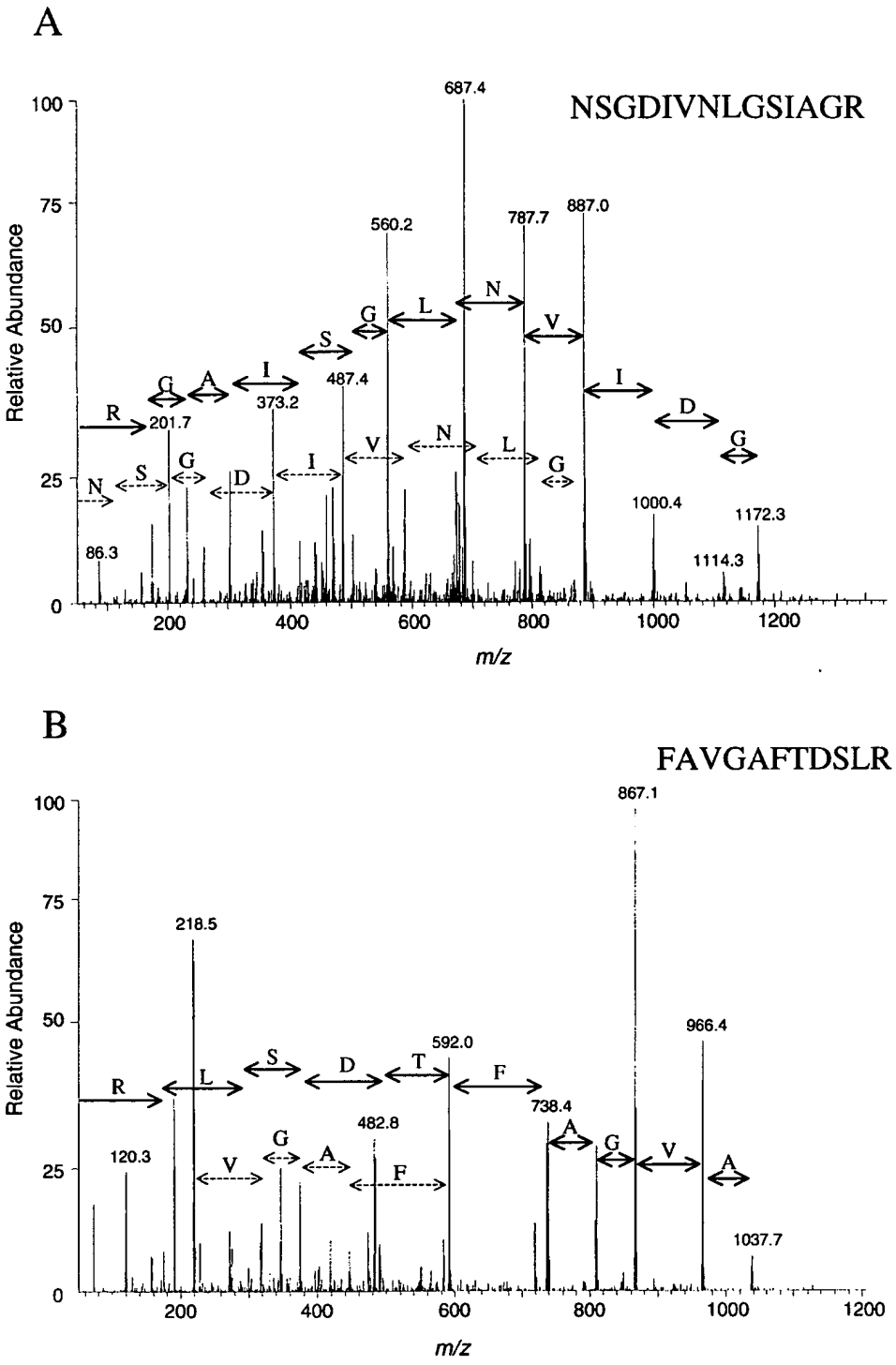


FIG. 3. Tandem mass (MS/MS) spectra resulting from analysis of a single spot on a 2D gel. The first quadrupole selected a single mass-to-charge ratio (*m/z*) of 687.2 (A) or 592.6 (B), while the collision cell was filled with argon gas, and a voltage which caused the peptide to undergo fragmentation by CID was applied. The third quadrupole scanned the mass range from 50 to 1,400 *m/z*. The computer program Sequest (8) was utilized to match MS/MS spectra to amino acid sequence by database searching. Both spectra matched peptides from the same protein, S57593 (yeast hypothetical protein YMR226C). Five other peptides from the same analysis were matched to the same protein.

migrating to the same gel spot, or (iv) the theoretical and observed pIs and molecular weights could not be reconciled. After these criteria were applied, the number of genes used in the correlation analysis was 106.

Codon bias and predicted half-lives. Codon bias is thought to be an indicator of protein expression, with highly expressed proteins having large codon bias values. The codon bias distribution for the entire set of more than 6,000 predicted yeast

gene ORFs is presented in Fig. 4A. The interval with the largest frequency of genes is between the codon bias values of 0.0 and 0.1. This segment contains more than 2,500 genes. The distribution of the codon bias values of the 128 different genes found in this study (all protein spots from Fig. 2) is shown in Fig. 4B, and protein half-lives (predicted from applying the N-end rule [33] to the experimentally determined or predicted protein N termini) are shown in Fig. 4C. No genes were identified with codon bias values less than 0.1 even though thousands of genes exist in this category. In addition, nearly all of the proteins identified had long predicted half-lives (greater than 30 h).

Correlation of mRNA and protein expression levels. The correlation between mRNA and protein levels of the genes selected as described above is shown in Fig. 5. For the entire group (106 genes) for which a complete data set was generated, there was a general trend of increased protein levels resulting from increased mRNA levels. The Pearson product moment correlation coefficient for the whole data set (106 genes) was 0.935. This number is highly biased by a small number of genes with very large protein and message levels. A more representative subset of the data is shown in the inset of Fig. 5. It shows genes for which the message level was below 10 copies/cell and includes 69% (73 of 106 genes) of the data used in the study. The Pearson product moment correlation coefficient for this data set was only 0.356. We also found that levels of protein expression coded for by mRNA with comparable abundance varied by as much as 30-fold and that the mRNA levels coding for proteins with comparable expression levels varied by as much as 20-fold.

The distortion of the correlation value induced by the uneven distribution of the data points along the x axis is further demonstrated by the analysis in Fig. 6. The 106 samples included in the study were ranked by protein abundance, and the Pearson product moment correlation coefficient was repeatedly calculated after including progressively more, and higher-abundance, proteins in each calculation. The correlation values remained relatively stable in the range of 0.1 to 0.4 if the lowest-expressed 40 to 95 proteins used in this study were included. However, the correlation value steadily climbed by the inclusion of each of the 11 very highly expressed proteins.

Correlation of protein and mRNA expression levels with codon bias. Codon bias is the propensity for a gene to utilize the same codon to encode an amino acid even though other codons would insert the identical amino acid in the growing polypeptide sequence. It is further thought that highly expressed proteins have large codon biases (3). To assess the value of codon bias for predicting mRNA and protein levels in exponentially growing yeast cells, we plotted the two experimental sets of data versus the codon bias (Fig. 7). The distribution patterns for both mRNA and protein levels with respect to codon bias were highly similar. There was high variability in the data within the codon bias range of 0.8 to 1.0. Although a large codon bias generally resulted in higher protein and message expression levels, codon bias did not appear to be predictive of either protein levels or mRNA levels in the cell.

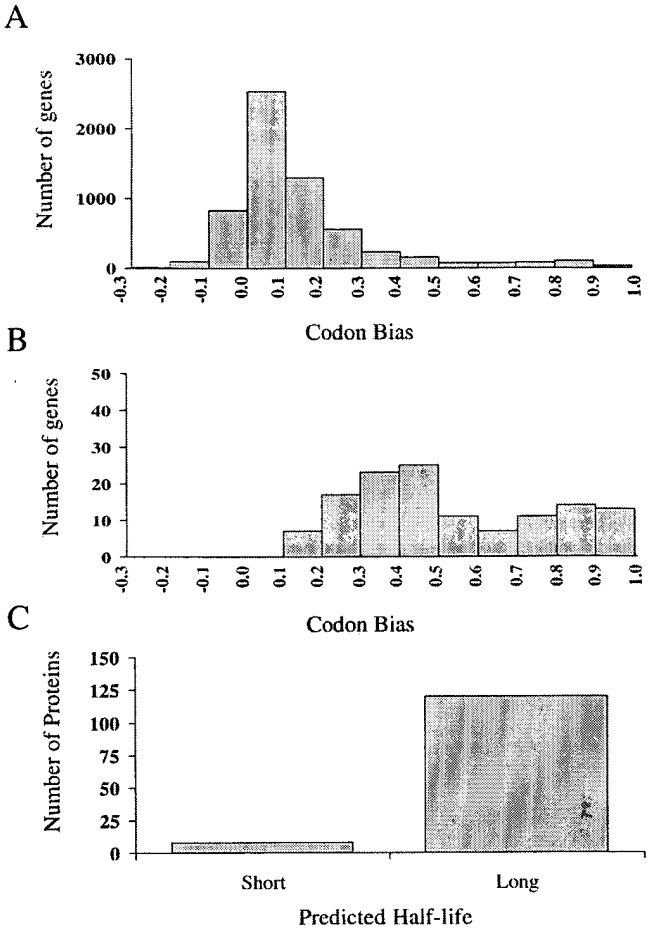


FIG. 4. Current proteome analysis technology utilizing 2DE without preenrichment samples mainly highly expressed and long-lived proteins. Genes encoding highly expressed proteins generally have large codon bias values. (A) Distribution of the yeast genome (more than 6,000 genes) based on codon bias. The interval with the largest frequency of genes is 0.0 to 0.1, with more than 2,500 genes. (B) Distribution of the genes from identified proteins in this study based on codon bias. No genes with codon bias values less than 0.1 were detected in this study. (C) Distribution of identified proteins in this study based on predicted half-life (estimated by N-end rule).

DISCUSSION

The desired end point for the description of a biological system is not the analysis of mRNA transcript levels alone but also the accurate measurement of protein expression levels and their respective activities. Quantitative analysis of global mRNA levels currently is a preferred method for the analysis of the state of cells and tissues (11). Several methods which either provide absolute mRNA abundance (34, 35) or relative mRNA levels in comparative analyses (20, 27) have been described elsewhere. The techniques are fast and exquisitely sensitive and can provide mRNA abundance for potentially any expressed gene. Measured mRNA levels are often implicitly or explicitly extrapolated to indicate the levels of activity of the corresponding protein in the cell. Quantitative analysis of protein expression levels (proteome analysis) is much more time-consuming because proteins are analyzed sequentially one by one and is not general because analyses are limited to the relatively highly expressed proteins. Proteome analysis does, however, provide types of data that are of critical importance for the description of the state of a biological system and that are not readily apparent from the sequence and the level of expression of the mRNA transcript. This study attempts to examine the relationship between mRNA and protein expression levels for a large number of expressed genes in cells representing the same state.

Limits in the sensitivity of current protein analysis technology precluded a completely random sampling of yeast proteins. We therefore based the study on those proteins visible by silver

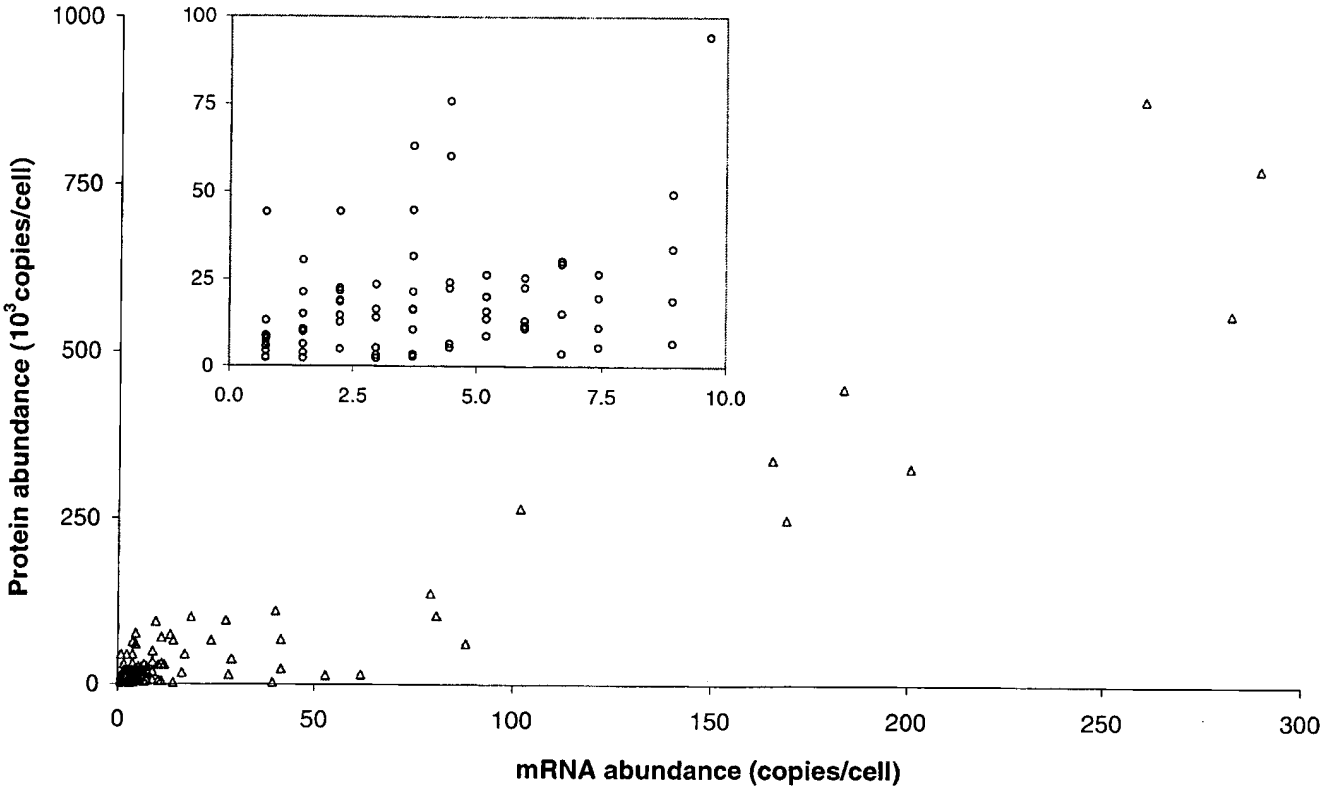


FIG. 5. Correlation between protein and mRNA levels for 106 genes in yeast growing at log phase with glucose as a carbon source. mRNA and protein levels were calculated as described in Materials and Methods. The data represent a population of genes with protein expression levels visible by silver staining on a 2D gel chosen to include the entire range of molecular weights, isoelectric focusing points, and staining intensities. The inset shows the low-end portion of the main figure. It contains 69% of the original data set. The Pearson product moment correlation for the entire data set was 0.935. The correlation for the inset containing 73 proteins (69%) was only 0.356.

staining on a 2D gel. Of the more than 1,000 visible spots, 156 were chosen to include the entire range of molecular weights, isoelectric focusing points, and staining intensities displayed on the 2D protein pattern. The genes identified in this study shared a number of properties. First, all of the proteins in this study had a codon bias of greater than 0.1 and 93% were greater than 0.2 (Fig. 4B). Second, with few exceptions, the proteins in this study had long predicted half-lives according to the N-end rule (Fig. 4C). Third, low-abundance proteins with regulatory functions such as transcription factors or protein kinases were not identified.

Because the population of proteins used in this study appears to be fairly homogeneous with respect to predicted half-life and codon bias, it might be expected that the correlation of the mRNA and protein expression levels would be stronger for this population than for a random sample of yeast proteins. We tested this assumption by evaluating the correlation value if different subsets of the available data were included in the calculation. The 106 proteins were ranked from lowest to highest protein expression level, and the trend in the correlation value was evaluated by progressively including more of the higher-abundance proteins in the calculation (Fig. 6). The correlation value when only the lower-abundance 40 to 93 proteins were examined was consistently between 0.1 and 0.4. If the 11 most abundant proteins were included, the correlation steadily increased to 0.94. We therefore expect that the correlation for all yeast proteins or for a random selection would be less than 0.4. The observed level of correlation between mRNA and protein expression levels suggests the importance of posttranslational mechanisms controlling gene expression. Such mechanisms include translational control (15) and control of protein half-life (33). Since these mechanisms are also active in higher eukaryotic cells, we speculate that there is no predictive correlation between steady-state levels of mRNA and those of protein in mammalian cells.

Like other large-scale analyses, the present study has several potential sources of error related to the methods used to determine mRNA and protein expression levels. The mRNA levels were calculated from frequency tables of SAGE data. This method is highly quantitative because it is based on actual sequencing of unique tags from each gene, and the number of times that a tag is represented is proportional to the number of mRNA molecules for a specific gene. This method has some limitations including the following: (i) the magnitude of the error in the measurement of mRNA levels is inversely proportional to the mRNA levels, (ii) SAGE tags from highly similar genes may not be distinguished and therefore are summed, (iii) some SAGE tags are from sequences in the 3′ untranslated region of the transcript, (iv) incomplete cleavage at the SAGE tag site by the restriction enzyme can result in two tags representing one mRNA, and (v) some transcripts actually do not generate a SAGE tag (34, 35).

For the SAGE method, the error associated with a value increases with a decreasing number of transcripts per cell. The conclusions drawn from this study are dependent on the quality of the mRNA levels from previously published data (35). Since more than 65% of the mRNA levels included in this study were calculated to 10 copies/cell or less (40% were less

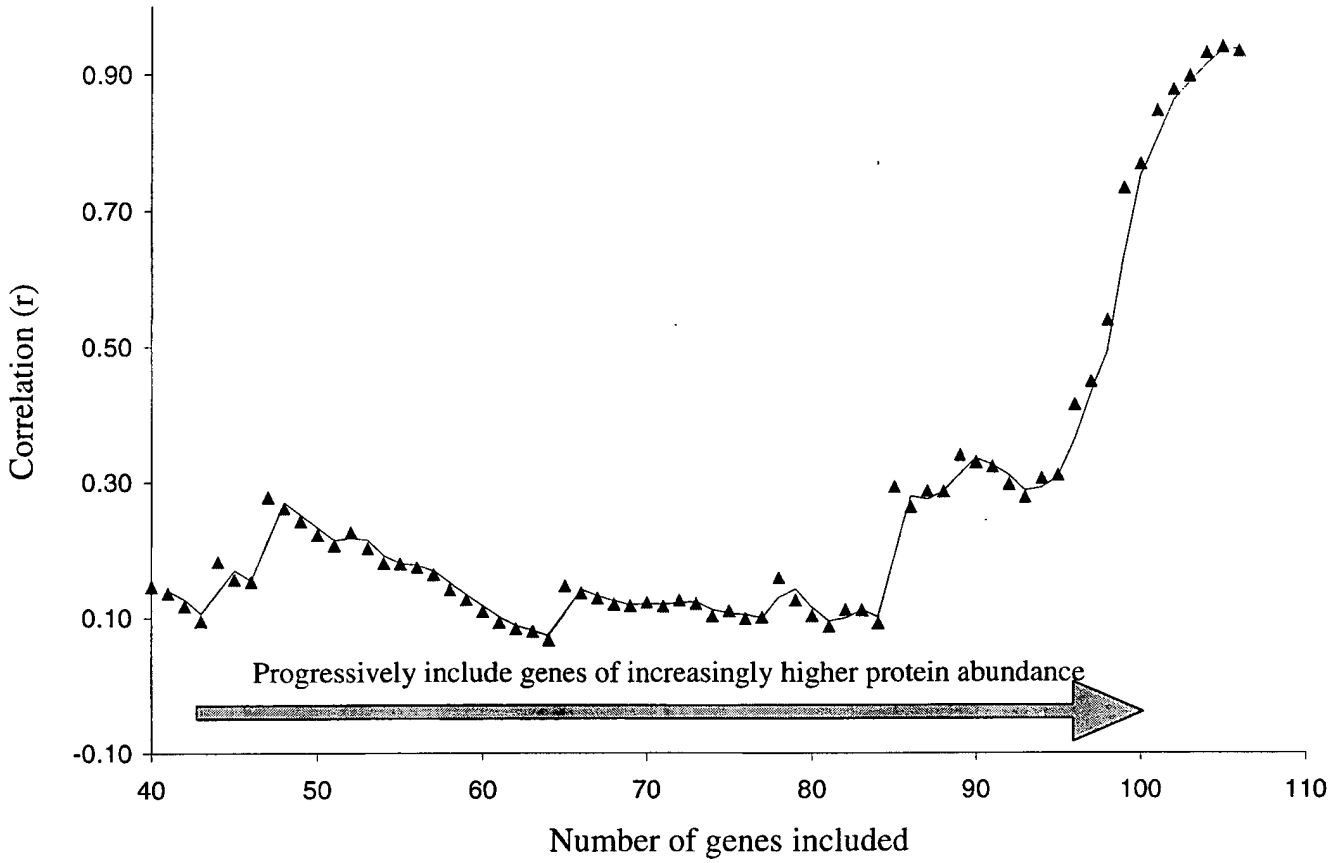


FIG. 6. Effect of highly abundant proteins on Pearson product moment correlation coefficient for mRNA and protein abundance in yeast. The set of 106 genes was ranked according to protein abundance, and the correlation value was calculated by including the 40 lowest-abundance genes and then progressively including the remaining 66 genes in order of abundance. The correlation value climbs as the final 11 highly abundant proteins are included.

than 4 copies/cell), the error associated with these values may be quite large. The mRNA levels were calculated from more than 20,000 transcripts. Assuming that the estimate of 15,000 mRNA molecules per cell is correct (16), this would mean that mRNA transcripts present at only a single copy per cell would be detected 72% of the time (35). The mRNA levels for each gene were carefully scrutinized, and only mRNA levels for which a high degree of confidence existed were included in the correlation value.

Protein abundance was determined by metabolic radiolabeling with [^{35}S]methionine. The calculation required knowledge of three variables: the number of methionines in the mature protein, the radioactivity contained in the protein, and the specific activity of the radiolabel normalized per methionine. The number of methionines per protein was determined from the amino acid sequence of the proteins identified by tandem mass spectrometry. For some proteins, it was not known whether the methionine of the nascent polypeptide was processed away. The N termini of those proteins were predicted based on the specificity of methionine aminopeptidase (31). If the N-terminal processing did not conform to the predicted specificity of processing enzymes, the calculation of the number of methionines would be affected. This discrepancy would affect most the quantitation of a protein with a very low number of methionines. The average number of calculated methionines per protein in this study was 7.2. We therefore expect the potential for erroneous protein quantitation due to unusual N-terminal processing to be small.

The amount of radioactivity contained in a single spot might be the sum of the radioactivity of comigrating proteins. Because protein identification was based on tandem mass spectrometric techniques, comigrating proteins could be identified. However, comigrating proteins were rarely detected in this study, most likely because relatively small amounts of total protein (40 μg) were initially loaded onto the gels, which resulted in highly focused spots containing generally 1 to 25 ng of protein. Because of the relatively small amount loaded, the concentrations of any potentially comigrating protein would likely be below the limit of detection of the mass spectrometry technique used in this study (1 to 5 ng) and below the limit of visualization by silver staining (1 to 5 ng). In the overwhelming majority of the samples analyzed, numerous peptides from a single protein were detected. It is assumed that any comigrating proteins were at levels too low to be detected and that their influence in the calculation would be small.

The specific activity of the radiolabel was determined by relating the precise amount of protein present in selected spots of a parallel gel, as determined by quantitative amino acid composition analysis, to the number of methionines present in the sequence of those proteins and the radioactivity determined by liquid scintillation counting. It is possible that the resulting number might be influenced by unavoidable losses inherent in the amino acid analysis procedure applied. Because four different proteins were utilized in the calculation and the experiment was done in duplicate, the specific activity calculated is thought to be highly accurate. Indeed, the specific

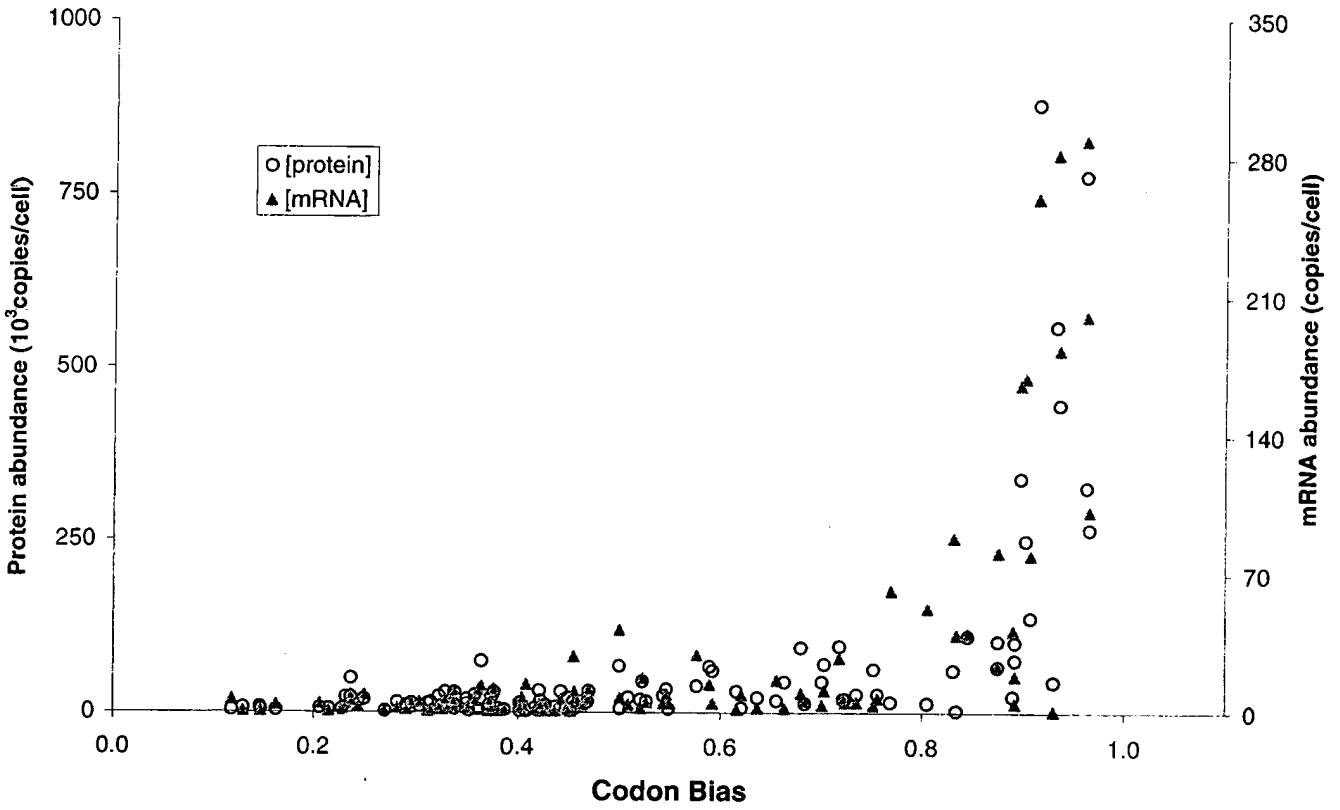


FIG. 7. Relationship between codon bias and protein and mRNA levels in this study. Yeast mRNA and protein expression levels were calculated as described in Materials and Methods. The data represent the same 106 genes as in Fig. 5.

activities calculated for each of the four proteins varied by less than 10%. Any inconsistencies in the calculation of the specific activity would result in differences in the absolute levels calculated but not in the relative numbers and would therefore not influence the correlation value determined.

The protein quantitative method used eliminates a number of potential errors inherent in previous methods for the quantitation of proteins separated by 2DE, such as preferential protein staining and bias caused by inequalities in the number of radiolabeled residues per protein. Any 2D gel-based method of quantitation is complicated by the fact that in some cases the translation products of the same mRNA migrated to different spots. One major reason is posttranslational modification or processing of the protein. Also, artifactual proteolysis during cell lysis and sample preparation can lead to multiple resolved forms of the protein. In such cases, the protein levels of spots coded for by the same mRNA were pooled. In addition, the existence of other spots coded for by the same mRNA that were not analyzed by mass spectrometry or that were below the limit of detection for silver staining cannot be ruled out. However, since this study is based on a class of highly expressed proteins, the presence of undetected minor spots below silver staining sensitivity corresponding to a protein analyzed in the study would generally cause a relatively small error in protein quantitation.

Codon bias is a measure of the propensity of an organism to selectively utilize certain codons which result in the incorporation of the same amino acid residue in a growing polypeptide chain. There are 61 possible codons that code for 20 amino acids. The larger the codon bias value, the smaller the number of codons that are used to encode the protein (19). It is thought that codon bias is a measure of protein abundance because highly expressed proteins generally have large codon bias values (3, 13).

Nearly all of the most highly expressed proteins had codon bias values of greater than 0.8. However, we detected a number of genes with high codon bias and relative low protein abundance (Fig. 7). For example, the expressed gene with both the second largest protein and mRNA levels in the study was ENO2_YEAST (775,000 and 289.1 copies/cell, respectively). ENO1_YEAST was also present in the gel at much lower protein and mRNA levels (44,200 and 0.7 copies/cell, respectively). The codon bias values for ENO2 and ENO1 are similar (0.96 and 0.93, respectively), but the expression of the two genes is differentially regulated. Specifically, ENO1_YEAST is glucose repressed (6) and was therefore present in low abundance under the conditions used. Other genes with large codon bias values that were not of high protein abundance in the gel include EFT1, TIF1, HXK2, GSP1, EGD2, SHM2, and TAL1. We conclude that merely determining the codon bias of a gene is not sufficient to predict its protein expression level.

Interestingly, codon bias appears to be an excellent indicator of the boundaries of current 2D gel proteome analysis technology. There are thousands of genes with expressed mRNA and likely expressed protein with codon bias values less than 0.1 (Fig. 4A). In this study, we detected none of them, and only a very small percentage of the genes detected in this study had codon bias values between 0.1 and 0.2 (Fig. 4B). Indeed, in every examined yeast proteome study (5, 7, 13, 28) where the combined total number of identified proteins is 300 to 400, this same observation is true. It is expected that for the more complex cells of higher eukaryotic organisms the detection of

low-abundance proteins would be even more challenging than for yeast. This indicates that highly abundant, long-lived proteins are overwhelmingly detected in proteome studies. If proteome analysis is to provide truly meaningful information about cellular processes, it must be able to penetrate to the level of regulatory proteins, including transcription factors and protein kinases. A promising approach is the use of narrow-range focusing gels with immobilized pH gradients (IPG) (23). This would allow for the loading of significantly more protein per pH unit covered and also provide increased resolution of proteins with similar electrophoretic mobilities. A standard pH gradient in an isoelectric focusing gel covers a 7-pH-unit range (pH 3 to 10) over 18 cm. A narrow-range focusing gel might expand the range to 0.5 pH units over 18 cm or more. This could potentially increase by more than 10-fold the number of proteins that can be detected. Clearly, current proteome technology is incapable of analyzing low-abundance regulatory proteins without employing an enrichment method for relatively low-abundance proteins. In conclusion, this study examined the relationship between yeast protein and message levels and revealed that transcript levels provide little predictive value with respect to the extent of protein expression.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation Science and Technology Center for Molecular Biotechnology, NIH grant T32HG00035-3, and a grant from Oxford Glycosciences.

We thank Jimmy Eng for expert computer programming, Garry Corthals and John R. Yates III for critical discussion, and Siavash Mohandesi for expert technical help.

REFERENCES

1. **Aebersold, R. H., J. Leavitt, R. A. Saavedra, L. E. Hood, and S. B. Kent.** 1987. Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis after in situ protease digestion on nitrocellulose. Proc. Natl. Acad. Sci. USA **84**:6970–6974.
2. **Aebersold, R. H., D. B. Teplow, L. E. Hood, and S. B. Kent.** 1986. Electroblotting onto activated glass. High efficiency preparation of proteins from analytical sodium dodecyl sulfate-polyacrylamide gels for direct sequence analysis. Eur. J. Biochem. **261**:4229–4238.
3. **Bennetzen, J. L., and B. D. Hall.** 1982. Codon selection in yeast. J. Biol. Chem. **257**:3026–3031.
4. **Boucherie, H., G. Dujardin, M. Kermorgant, C. Monribot, P. Slonimski, and M. Perrot.** 1995. Two-dimensional protein map of Saccharomyces cerevisiae: construction of a gene-protein index. Yeast **11**:601–613.
5. **Boucherie, H., F. Sagliocco, R. Joubert, I. Maillet, J. Labarre, and M. Perrot.** 1996. Two-dimensional gel protein database of Saccharomyces cerevisiae. Electrophoresis **17**:1683–1699.
6. **Carmen, A. A., P. K. Brindle, C. S. Park, and M. J. Holland.** 1995. Transcriptional regulation by an upstream repression sequence from the yeast enolase gene ENO1. Yeast **11**:1031–1043.
7. **Ducret, A., I. VanOostveen, J. K. Eng, J. R. Yates, and R. Aebersold.** 1998. High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry. Protein Sci. **7**:706–719.
8. **Eng, J., A. McCormack, and J. R. Yates.** 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. **5**:976–989.
9. **Figeys, D., A. Ducret, J. R. Yates, and R. Aebersold.** 1996. Protein identification by solid phase microextraction-capillary zone electrophoresis-microelectrospray-tandem mass spectrometry. Nat. Biotechnol. **14**:1579–1583.
10. **Figeys, D., I. VanOostveen, A. Ducret, and R. Aebersold.** 1996. Protein identification by capillary zone electrophoresis/microelectrospray ionization-tandem mass spectrometry at the subfemtomole level. Anal. Chem. **68**:1822–1828.
11. **Fraser, C. M., and R. D. Fleischmann.** 1997. Strategies for whole microbial genome sequencing and analysis. Electrophoresis **18**:1207–1216.
12. **Garrels, J. I., B. Futcher, R. Kobayashi, G. I. Latter, B. Schwender, T. Volpe, J. R. Warner, and C. S. McLaughlin.** 1994. Protein identifications for a Saccharomyces cerevisiae protein database. Electrophoresis **15**:1466–1486.
13. **Garrels, J. I., C. S. McLaughlin, J. R. Warner, B. Futcher, G. I. Latter, R. Kobayashi, B. Schwender, T. Volpe, D. S. Anderson, F. Mesquita-Fuentes, and W. E. Payne.** 1997. Proteome studies of Saccharomyces cerevisiae: identification and characterization of abundant proteins. Electrophoresis **18**:1347–1360.
14. **Gygi, S. P., and R. Aebersold.** 1998. Absolute quantitation of 2-DE protein spots, p. 417–421. *In* A. J. Link (ed.), 2-D protocols for proteome analysis. Humana Press, Totowa, N.J.
15. **Harford, J. B., and D. R. Morris.** 1997. Post-transcriptional gene regulation. Wiley-Liss, Inc., New York, N.Y.
16. **Hereford, L. M., and M. Rosbash.** 1977. Number and distribution of polyadenylated RNA sequences in yeast. Cell **10**:453–462.
17. **Hodges, P. E., W. E. Payne, and J. I. Garrels.** 1998. The Yeast Protein Database (YPD): a curated proteome database for Saccharomyces cerevisiae. Nucleic Acids Res. **26**:68–72.
18. **Klose, J., and U. Kobalz.** 1995. Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome. Electrophoresis **16**:1034–1059.
19. **Kurland, C. G. 1991.** Codon bias and gene expression. FEBS Lett. **285**:165–169.
20. **Lashkari, D. A., J. L. DeRisi, J. H. McCusker, A. F. Namath, C. Gentile, S. Y. Hwang, P. O. Brown, and R. W. Davis.** 1997. Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc. Natl. Acad. Sci. USA **94**:13057–13062.
21. **Liang, P., and A. B. Pardee.** 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science **257**:967–971.
22. **Link, A. J., L. G. Hays, E. B. Carmack, and J. R. Yates III.** 1997. Identifying the major proteome components of Haemophilus influenzae type-strain NCTC 8143. Electrophoresis **18**:1314–1334.
23. **Nawrocki, A., M. R. Larsen, A. V. Podtelejnikov, O. N. Jensen, M. Mann, P. Roepstorff, A. Gorg, S. J. Fey, and P. M. Larsen.** 1998. Correlation of acidic and basic carrier ampholyte and immobilized pH gradient two-dimensional gel electrophoresis patterns based on mass spectrometric protein identification. Electrophoresis **19**:1024–1035.
24. **O'Farrell, P. H. 1975.** High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. **250**:4007–4021.

24a. **OWL Protein Sequence Database.** 2 August 1998, posting date. [Online.] http://bmbsgi11.leeds.ac.uk/bmb5dp/owl.html. [8 January 1999, last date accessed.]

25. **Patterson, S. D., and R. Aebersold.** 1995. Mass spectrometric approaches for the identification of gel-separated proteins. Electrophoresis **16**:1791–1814.
26. **Pennington, S. R., M. R. Wilkins, D. F. Hochstrasser, and M. J. Dunn.** 1997. Proteome analysis: from protein characterization to biological function. Trends Cell Biol. **7**:168–173.
27. **Shalon, D., S. J. Smith, and P. O. Brown.** 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res. **6**:639–645.
28. **Shevchenko, A., O. N. Jensen, A. V. Podtelejnikov, F. Sagliocco, M. Wilm, O. Vorm, P. Mortensen, H. Boucherie, and M. Mann.** 1996. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. Proc. Natl. Acad. Sci. USA **93**:14440–14445.
29. **Shevchenko, A., M. Wilm, O. Vorm, and M. Mann.** 1996. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. Anal. Chem. **68**:850–858.
30. **Sikorski, R. S., and P. Hieter.** 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics **122**:19–27.
31. **Tsunasawa, S., J. W. Stewart, and F. Sherman.** 1985. Amino-terminal processing of mutant forms of yeast iso-1-cytochrome c. The specificities of methionine aminopeptidase and acetyltransferase. J. Biol. Chem. **260**:5382–5391.
32. **Urlinger, S., K. Kuchler, T. H. Meyer, S. Uebel, and R. Tamp'e.** 1997. Intracellular location, complex formation, and function of the transporter associated with antigen processing in yeast. Eur. J. Biochem. **245**:266–272.
33. **Varshavsky, A.** 1996. The N-end rule: functions, mysteries, uses. Proc. Natl. Acad. Sci. USA **93**:12142–12149.
34. **Velculescu, V. E., L. Zhang, B. Vogelstein, and K. W. Kinzler.** 1995. Serial analysis of gene expression. Science **270**:484–487.
35. **Velculescu, V. E., L. Zhang, W. Zhou, J. Vogelstein, M. A. Basrai, D. E. Bassett, Jr., P. Hieter, B. Vogelstein, and K. W. Kinzler.** 1997. Characterization of the yeast transcriptome. Cell **88**:243–251.
36. **Wilkins, M. R., K. L. Williams, R. D. Appel, and D. F. Hochstrasser.** 1997. Proteome research: new frontiers in functional genomics. Springer-Verlag, Berlin, Germany.
37. **Wilm, M., A. Shevchenko, T. Houthaeve, S. Breit, L. Schweigerer, T. Fotsis, and M. Mann.** 1996. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. Nature **379**:466–469.
38. **Yan, J. X., M. R. Wilkins, K. Ou, A. A. Gooley, K. L. Williams, J. C. Sanchez, O. Golaz, C. Pasquali, and D. F. Hochstrasser.** 1996. Large-scale amino-acid analysis for proteome studies. J. Chromatogr. A **736**:291–302.
39. **YPD Website.** 6 March 1998, revision date. [Online.] Proteome, Inc. http://www.proteome.com/YPDhome.html. [8 January 1999, last date accessed.]

editorial Journal of proteome research

Do We Have Enough Biomarkers?

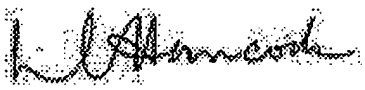
The Editor has become aware of a recent push to validate currently available biomarkers in an extensive clinical setting. The reasoning behind such a push is that there are already a significant number of biomarkers that now need to be used effectively in the clinic. Many biomarkers, such as the carcinoembryonic antigen, have been known for some time and are used widely for patient management. The older biomarkers, however, are not effective for early diagnosis.

With the advent of genomics and, later, proteomics, there has been a substantial investment in using these new tools to generate additional biomarkers. The problem with this new information is that it is too early to get consensus on what is a useful marker or what is a good patient population for such a study. Therefore, it is unclear whether the new markers currently in hand will give better clinical information than the ones that have been used in the past. An additional problem is that the markers that are generated by proteomics are not always consistent with the markers that are generated from expression profiling.

The challenge in this situation is to balance the need of patients for better, early diagnosis of disease with the need to have high-quality markers for the expensive and time-consuming validation process. This Editor believes that proteomics is at too early a stage for this new technology to have generated a quality list of markers. The risk is if we push the existing markers into extensive clinical validation, we will be missing the fruits of improvements in emerging proteomics technology. I think many people in the proteomics community would agree that federal granting agencies should be enticed to continue investments in basic proteomics technology. In addition, funding should be made available for basic science studies that will continue to generate biomarkers, and there needs to be some type of consensus-building process that can lead to a consolidation of the different lists of biomarkers.

There are good past models for such activities, such as the consensus-forming meetings that the U.S. Food and Drug Administration has held; these yielded technical innovations. One example was the generation of new protein pharmaceuticals at the advent of the biotechnology industry. Another example, in the early days of the genome sequencing program, was when a group of experts came together to agree on annotation of the early results. The Human Proteome Organization is a good example of an international group of laboratories coming together to consolidate the output from a number of studies with different technology platforms.

I would like to encourage the biomedical community not to rush to judgment in terms of biomarkers, but instead to give research more time to produce quality biomarker information. Then we should conduct a thorough evaluation of a widely agreed-on list before we attempt to determine which of these new markers are indeed worthy of extensive clinical investigation.



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	:	Eaton, et al.
Appl. No.	:	10/063,557
Filed	:	May 2, 2002
For	:	SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
Examiner	:	David J. Blanchard
Group Art Unit	:	1642

DECLARATION OF J. CHRISTOPHER GRIMALDI, UNDER 37 CFR §1.132

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

I, J. Christopher Grimaldi, declare and state as follows:

1. I am a Senior Research Associate in the Molecular Biology Department of Genentech, Inc., South San Francisco, CA 94080.

2. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).

3. I joined Genentech in January of 1999. From 1999 to 2003, I directed the Cloning Laboratory in the Molecular Biology Department. During this time I directed or performed numerous molecular biology techniques including semi-quantitative Polymerase Chain Reaction (PCR) analyses. I am currently involved, among other projects, in the isolation of genes coding for membrane associated proteins which can be used as targets for antibody therapeutics against cancer. In connection with the above-identified patent application, I personally performed or directed the semi-quantitative PCR gene expression analyses in the assay entitled "Tumor Versus Normal Differential Tissue Expression Distribution," which is described in EXAMPLE 18 in the specification. These studies were used to identify differences in gene expression between tumor tissue and their normal counterparts.

4. EXAMPLE 18 reports the results of the PCR analyses conducted as part of the investigating of several newly discovered DNA sequences. This process included developing

primers and analyzing expression of the DNA sequences of interest in normal and tumor tissues. The analyses were designed to determine whether a difference exists between gene expression in normal tissues as compared to tumor in the same tissue type.

5. The DNA libraries used in the gene expression studies were made from pooled samples of normal and of tumor tissues. Data from pooled samples is more likely to be accurate than data obtained from a sample from a single individual. That is, the detection of variations in gene expression is likely to represent a more generally relevant condition when pooled samples from normal tissues are compared with pooled samples from tumors in the same tissue type.

6. In differential gene expression studies, one looks for genes whose expression levels differ significantly under different conditions, for example, in normal versus diseased tissue. Thus, I conducted a semi-quantitative analysis of the expression of the DNA sequences of interest in normal versus tumor tissues. Expression levels were graded according to a scale of +, -, and +/- to indicate the amount of the specific signal detected. Using the widely accepted technique of PCR, it was determined whether the polynucleotides tested were more highly expressed, less expressed, or whether expression remained the same in tumor tissue as compared to its normal counterpart. Because this technique relies on the visual detection of ethidium bromide staining of PCR products on agarose gels, it is reasonable to assume that any detectable differences seen between two samples will represent at least a two fold difference in cDNA.

7. The results of the gene expression studies indicate that the genes of interest can be used to differentiate tumor from normal. The precise levels of gene expression are irrelevant; what matters is that there is a relative difference in expression between normal tissue and tumor tissue. The precise type of tumor is also irrelevant; again, the assay was designed to indicate whether a difference exists between normal tissue and tumor tissue of the same type. If a difference is detected, this indicates that the gene and its corresponding polypeptide and antibodies against the polypeptide are useful for diagnostic purposes, to screen samples to differentiate between normal and tumor. Additional studies can then be conducted if further information is desired.

8. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

By: _____ Date: 8/10/2004

J. Christopher Grimaldi

S:\DOCS\AOK\AOK-5591.DOC/081004

J. Christopher Grimaldi

1434-36th Ave.
San Francisco, CA 94122
(415) 681-1639 (Home)

EDUCATION University of California, Berkeley
Bachelor of Arts in Molecular Biology, 1984

EMPLOYMENT EXPERIENCE

SRA Genentech Inc., South San Francisco; 1/99 to present

Previously, was responsible to direct and manage the Cloning Lab. Currently focused on isolating cancer specific genes for the Tumor Antigen (TAP), and Secreted Tumor Protein (STOP) projects for the Oncology Department as well as Immunologically relevant genes for the Immunology Department. Directed a lab of 6 scientists focused on a company-wide team effort to identify and isolate secreted proteins for potential therapeutic use (SPDI). For the SPDI project my duties were, among other things, the critically important coordination of the cloning of thousands of putative genes, by developing a smooth process of communication between the Bioinformatics, Cloning, Sequencing, and Legal teams. Collaborated with several groups to discover novel genes through the Curagen project, a unique differential display methodology. Interacted extensively with the Legal team providing essential data needed for filing patents on novel genes discovered through the SPDI, TAP and Curagen projects. My group has developed, implemented and patented high throughput cloning methodologies that have proven to be essential for the isolation of hundreds of novel genes for the SPDI, TAP and Curagen projects as well as dozens of other smaller projects.

Scientist DNAX Research Institute, Palo Alto; 9/91 to 1/99

Involved in multiple projects aimed at understanding novel genes discovered through bioinformatics studies and functional assays. Developed and patented a method for the specific depletion of eosinophils in vivo using monoclonal antibodies. Developed and implemented essential technical methodologies and provided strategic direction in the areas of expression, cloning, protein purification, general molecular biology, and monoclonal antibody production. Trained and supervised numerous technical staff.

Facilities Manager Corixa, Redwood City; 5/89 - 7/91.

Directed plant-related activities, which included expansion planning, maintenance, safety, purchasing, inventory control, shipping and receiving, and laboratory management. Designed and implemented the safety program. Also served as liaison to regulatory agencies at the local, state and federal level. Was in charge of property leases, leasehold improvements, etc. Negotiated vendor contracts and directed the purchasing department. Trained and supervised personnel to carry out the above-mentioned duties.

SRA University of California, San Francisco
Cancer Research Institute; 2/87-4/89.

Was responsible for numerous cloning projects including: studies of somatic hypermutation, studies of AIDS-associated lymphomas, and cloning of t(5;14), t(11;14), and t(8;14) translocations. Focused on the activation of hemopoietic growth factors involved in the t(5;14) translocation in leukemia patients..

Research
Technician Berlex Biosciences, South San Francisco; 7/85-2/87.

Worked on a subunit porcine vaccine directed against Mycoplasma hyopneumoniae. Was responsible for generating genomic libraries, screening with degenerate oligonucleotides, and characterizing and expressing clones in E. coli. Also constructed a general purpose expression vector for use by other scientific teams.

PUBLICATIONS

1. Hilary F. Clark, et al. "The Secreted Protein Discovery Initiative (SPDI), a Large-scale Effort to Identify Novel Human Secreted and Transmembrane Proteins: a bioinformatics assessment." Genome Res. Vol 13(10), 2265-2270, 2003

2. Sean H. Adams, Clarissa Chui' Sarah L. Schilbach, Xing Xian Yu, Audrey D. Goddard, J. Christopher Grimaldi, James Lee, Patrick Dowd, David A. Lewin, & Steven Colman' "BFIT, a Unique Acyl-CoA Thioesterase Induced in Thermogenic Brown Adiopose Tissue: Cloning, organization of the humanb gene and assessment of a potential link to obesity" Biochemical Journal, Vol 360, 135-142, 2001

3. Szeto W, Jiang W, Tice DA, Rubinfeld B, Hollingshead PG, Fong SE, Dugger DL, Pham T, Yansura D, Wong TA, Grimaldi JC, Corpuz RT, Singh JS, Frantz GD, , Devaux B, Crowley CW, Schwall RH, Eberhard DA, Rastelli L, Polakis.P, and Pennica D. "Overexpression of the Retenoic Acid-Responsive Gene Stra6 in Human Cancers and its Synergistic Activation by Wnt-1 and Retinoic Acid." Cancer Research Vol. 61(10), 4197-4205, 2001

4. Jeanne Kahn, Fuad Mehraban, Gladdys Ingle, Xiaohua Xin, Juliet E. Bryant, Gordon Vehar, Jill Schoenfeld, J. Christopher Grimaldi (incorrectly named as "Grimaldi, CJ"), Franklin Peale, Aparna Draksharapu, David A. Lewin, and Mary E. Gerritsen. "Gene Expression Profiling in an in Vitro Model of Angiogenesis." American Journal of Pathology Vol 156(6), 1887-1900, 2000.

5. Grimaldi JC, Yu NX, Grunig G, Seymour BW, Cottrez F, Robinson DS, Hosken N, Ferlin WG, Wu X, Soto H, O'Garra A, Howard MC, Coffman RL. "Depletion of eosinophils in mice through the use of antibodies specific for C-C chemokine receptor 3 (CCR3). Journal of Leukocyte Biology; Vol. 65(6), 846-53, 1999

6. Oliver AM, Grimaldi JC, Howard MC, Kearney JF. "Independently ligating CD38 and Fc gammaRIIB relays a dominant negative signal to B cells." Hybridoma Vol. 18(2), 113-9, 1999

7. Cockayne DA, Muchamuel T, Grimaldi JC, Muller-Steffner H, Randall TD, Lund FE, Murray R, Schuber F, Howard MC. "Mice deficient for the ecto-nicotinamide adenine dinucleotide glycohydrolase CD38 exhibit altered humoral immune responses." Blood Vol. 92(4), 1324-33, 1998

8. Frances E. Lund, Nanette W. Solvason, Michael P. Cooke, Andrew W. Heath, J. Christopher Grimaldi, Troy D. Randall, R. M. E. Parkhouse, Christopher C Goodnow and Maureen C. Howard. "Signaling through murine CD38 is impaired in antigen receptor unresponsive B cells." European Journal of Immunology, Vol. 25(5), 1338-1345, 1995

9. M. J. Guimaraes, J. F. Bazan, A. Zolotnik, M. V. Wiles, J. C. Grimaldi, F. Lee, T. McClanahan. "A new approach to the study of haematopoietic development in the yolk sac and embryoid body." Development, Vol. 121(10), 3335-3346, 1995

10. J. Christopher Grimaldi, Sriram Balasubramanian, J. Fernando Bazan, Armen Shanafelt, Gerard Zurawski and Maureen Howard. "CD38-mediated protein ribosylation." Journal of Immunology,Vol. 155(2), 811-817, 1995

11. Leopoldo Santos-Argumedo, Frances F. Lund, Andrew W. Heath, Nanette Solvason, Wei Wei Wu, J. Christopher Grimaldi, R. M. E. Parkhouse and Maureen Howard. "CD38 unresponsiveness of xid B cells implicates Bruton's tyrosine kinase (btk) as a regulator of CD38 induced signal transduction." International Immunology, Vol 7(2), 163-170, 1995

12. Frances Lund, Nanette Solvason, J. Christopher Grimaldi, R. M. E. Parkhouse and Maureen Howard. "Murine CD38: An immunoregulatory ectoenzyme." Immunology Today, Vol. 16(10), 469-473, 1995

13. Maureen Howard, J. Christopher Grimaldi, J. Fernando Bazan, Frances E. Lund, Leopoldo Santos-Argumedo, R. M. E. Parkhouse, Timothy F. Walseth, and Hon Cheung Lee. "Formation and Hydrolysis of Cyclic ADP-Ribose Catalyzed by Lymphocyte Antigen CD38." Science, Vol. 262, 1056-1059, 1993

14. Nobuyuki Harada, Leopoldo Santos-Argumedo, Ray Chang, J. Christopher Grimaldi, Frances Lund, Camilynn I. Brannan, Neal G. Copeland, Nancy A. Jenkins, Andrew Heath, R. M. E. Parkhouse and Maureen Howard. "Expression Cloning of a cDNA Encoding a Novel Murine B Cell Activation Marker: Homology to Human CD38." The Journal of Immunology, Vol. 151, 3111-3118, 1993

15. David J. Rawlings, Douglas C. Saffran, Satoshi Tsukada, David A. Largaespada, J. Christopher Grimaldi, Lucie Cohen Randolph N. Mohr, J. Fernando Bazan, Maureen Howard, Neal G. Copeland, Nancy A. Jenkins, Owen Witte. "Mutation of Unique Region of Bruton's Tyrosine Kinase in Immunodeficient XID Mice." Science, Vol. 261, 358-360, 1993

16. J. Christopher Grimaldi, Raul Torres, Christine A. Kozak, Ray Chang, Edward Clark, Maureen Howard, and Debra A. Cockayne. "Genomic Structure and Chromosomal Mapping of the Murine CD40 Gene." The Journal of Immunology, Vol 149, 3921-3926, 1992

17. Timothy C. Meeker, Bruce Shiramizu, Lawrence Kaplan, Brian Herndier, Henry Sanchez, J. Christopher Grimaldi, James Baumgartner, Jacab Rachlin, Ellen Feigal, Mark Rosenblum and Michael S. McGrath. "Evidence for Molecular Subtypes of HIV-Associated Lymphoma:

Division into Peripheral Monoclonal, Polyclonal and Central Nervous System Lymphoma." AIDS, Vol. 5, 669-674, 1991

18. Ann Grimaldi and Chris Grimaldi. "Small-Scale Lambda DNA Prep." Contribution to Current Protocols in Molecular Biology, Supplement 5, Winter 1989

19. J. Christopher Grimaldi, Timothy C. Meeker. "The t(5;14) Chromosomal Translocation in a Case of Acute Lymphocytic Leukemia Joins the Interleukin-3 Gene to the Immunoglobulin Heavy Chain Gene." Blood, Vol. 73, 2081-2085, 1989

20. Timothy C. Meeker, J. Christopher Grimaldi, et al. "An Additional Breakpoint Region in the BCL-1 Locus Associated with the t(11;14) (q13;q32) Translocation of B-Lymphocytic Malignancy." Blood, Vol. 74, 1801-1806, 1989

21 Timothy C. Meeker, J. Christopher Grimaldi, Robert O'Rourke, et al. "Lack of Detectable Somatic Hypermutation in the V Region of the Ig H Chain Gene of a Human Chronic B Lymphocytic Leukemia." The Journal of Immunology, Vol. 141, 3994-3998, 1988

MANUSCRIPTS IN PREPARATION

1. Sriram Balasubramanian, J. Christopher Grimaldi, J. Fernando Bazan, Gerard Zurawski and Maureen Howard. "Structural and functional characterization of CD38: Identification of active site residues"

PATENTS

1. "Methods for Eosinophil Depletion with Antibody to CCR3 Receptor" (US 6,207,155 B1).

2. "Amplification Based Cloning Method." (US 6,607,899)

3. Ashkenazi et al., "Secreted and Transmembrane Polypeptides and Nucleic Acids Encoding the Same." (this patent covers several hundred genes)

4. "IL-17 Homologous Polypeptides and Therapeutic Uses Thereof"

5. "Method of Diagnosing and Treating Cartilaginous Disorders."

MEMBERSHIPS AND ACTIVITIES

Editor	Frontiers in Bioscience
Member	DNAX Safety Committee 1991-1999 Biological Safety Affairs Forum (BSAF) 1990-1991 Environmental Law Foundation (ELF) 1990-1991

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	:	Eaton, et al.
Appl. No.	:	10/063,557
Filed	:	May 2, 2002
For	:	SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
Examiner	:	David J. Blanchard
Group Art Unit	:	1642

DECLARATION OF J. CHRISTOPHER GRIMALDI, UNDER 37 C.F.R. § 1.132

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

I, J. Christopher Grimaldi, declare and say as follows:

1. I am a Senior Research Associate in the Molecular Biology Department of Genentech, Inc., South San Francisco, CA 94080.

2. I joined Genentech in January of 1999. From 1999 to 2003, I directed the Cloning Laboratory in the Molecular Biology Department. During this time I directed or performed numerous molecular biology techniques including qualitative Polymerase Chain Reaction (PCR) analyses. I am currently involved in, among other projects, the isolation of genes coding for membrane associated proteins which can be used as targets for antibody therapeutics against cancer. In connection with the above-identified patent application, I personally performed or directed the semi-quantitative PCR analyses in the assay entitled "Tumor Versus Normal Differential Tissue Expression Distribution" which is described in EXAMPLE 18 in the specification that were used to identify differences in gene expression between tumor tissue and their normal counterparts.

3. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).

4. In differential gene expression studies, one looks for genes whose expression levels differ significantly under different conditions, for example, in normal versus diseased tissue.

Chromosomal aberrations, such as gene amplification, and chromosomal translocations are important markers of specific types of cancer and lead to the aberrant expression of specific genes and their encoded polypeptides, including over-expression and under-expression. For example, gene amplification is a process in which specific regions of a chromosome are duplicated, thus creating multiple copies of certain genes that normally exist as a single copy. Gene under-expression can occur when a gene is not transcribed into mRNA. In addition, chromosomal translocations occur when two different chromosomes break and are rejoined to each other chromosome resulting in a chimeric chromosome which displays a different expression pattern relative to the parent chromosomes. Amplification of certain genes such as Her2/Neu [Singleton *et al.*, Pathol. Annu., 27Pt1:165-190], or chromosomal translocations such as t(5;14), [Grimaldi *et al.*, Blood, 73(8):2081-2085(1989); Meeker *et al.*, Blood, 76(2):285-289(1990)] give cancer cells a growth or survival advantage relative to normal cells, and might also provide a mechanism of tumor cell resistance to chemotherapy or radiotherapy. When the chromosomal aberration results in the aberrant expression of a mRNA and the corresponding gene product (the polypeptide), as it does in the aforementioned cases, the gene product is a promising target for cancer therapy, for example, by the therapeutic antibody approach.

5. Comparison of gene expression levels in normal versus diseased tissue has important implications both diagnostically and therapeutically. For example, those who work in this field are well aware that in the vast majority of cases, when a gene is over-expressed, as evidenced by an increased production of mRNA, the gene product or polypeptide will also be over-expressed. It is unlikely that one identifies increased mRNA expression without associated increased protein expression. This same principle applies to gene under-expression. When a gene is under-expressed, the gene product is also likely to be under-expressed. Stated in another way, two cell samples which have differing mRNA concentrations for a specific gene are expected to have correspondingly different concentration of protein for that gene. Techniques used to detect mRNA, such as Northern Blotting, Differential Display, *in situ* hybridization, quantitative PCR, Taqman, and more recently Microarray technology all rely on the dogma that a change in mRNA will represent a similar change in protein. If this dogma did not hold true then these techniques would have little value and not be so widely used. The use of mRNA quantitation techniques have identified a seemingly endless number of genes which are differentially expressed in various tissues and these genes have subsequently been shown to have correspondingly similar changes in their protein levels. Thus, the detection of increased mRNA expression is expected to result in increased polypeptide expression, and the detection of decreased mRNA expression is expected to result in decreased polypeptide expression. The detection of increased or decreased polypeptide expression can be used for cancer diagnosis and treatment.

6. However, even in the rare case where the protein expression does not correlate with the mRNA expression, this still provides significant information useful for cancer diagnosis and treatment. For example, if over- or under-expression of a gene product does not correlate with over- or under-expression of mRNA in certain tumor types but does so in others, then identification of both gene expression and protein expression enables more accurate tumor classification and hence better determination of suitable therapy. In addition, absence of over- or

under-expression of the gene product in the presence of a particular over- or under-expression of mRNA is crucial information for the practicing clinician. For example, if a gene is over-expressed but the corresponding gene product is not significantly over-expressed, the clinician accordingly will decide not to treat a patient with agents that target that gene product.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

By: _____ Date: 8/10/2004

J. Christopher Grimaldi

S:\DOCS\AOK\AOK-5479.DOC
071904

J. Christopher Grimaldi

1434-36th Ave.
San Francisco, CA 94122
(415) 681-1639 (Home)

EDUCATION University of California, Berkeley
Bachelor of Arts in Molecular Biology, 1984

EMPLOYMENT EXPERIENCE

SRA Genentech Inc., South San Francisco; 1/99 to present

Previously, was responsible to direct and manage the Cloning Lab. Currently focused on isolating cancer specific genes for the Tumor Antigen (TAP), and Secreted Tumor Protein (STOP) projects for the Oncology Department as well as Immunologically relevant genes for the Immunology Department. Directed a lab of 6 scientists focused on a company-wide team effort to identify and isolate secreted proteins for potential therapeutic use (SPDI). For the SPDI project my duties were, among other things, the critically important coordination of the cloning of thousands of putative genes, by developing a smooth process of communication between the Bioinformatics, Cloning, Sequencing, and Legal teams. Collaborated with several groups to discover novel genes through the Curagen project, a unique differential display methodology. Interacted extensively with the Legal team providing essential data needed for filing patents on novel genes discovered through the SPDI, TAP and Curagen projects. My group has developed, implemented and patented high throughput cloning methodologies that have proven to be essential for the isolation of hundreds of novel genes for the SPDI, TAP and Curagen projects as well as dozens of other smaller projects.

Scientist DNAX Research Institute, Palo Alto; 9/91 to 1/99

Involved in multiple projects aimed at understanding novel genes discovered through bioinformatics studies and functional assays. Developed and patented a method for the specific depletion of eosinophils in vivo using monoclonal antibodies. Developed and implemented essential technical methodologies and provided strategic direction in the areas of expression, cloning, protein purification, general molecular biology, and monoclonal antibody production. Trained and supervised numerous technical staff.

Facilities Manager Corixa, Redwood City; 5/89 - 7/91.

Directed plant-related activities, which included expansion planning, maintenance, safety, purchasing, inventory control, shipping and receiving, and laboratory management. Designed and implemented the safety program. Also served as liaison to regulatory agencies at the local, state and federal level. Was in charge of property leases, leasehold improvements, etc. Negotiated vendor contracts and directed the purchasing department. Trained and supervised personnel to carry out the above-mentioned duties.

SRA University of California, San Francisco
Cancer Research Institute; 2/87-4/89.

Was responsible for numerous cloning projects including: studies of somatic hypermutation, studies of AIDS-associated lymphomas, and cloning of t(5;14), t(11;14), and t(8;14) translocations. Focused on the activation of hemopoietic growth factors involved in the t(5;14) translocation in leukemia patients..

Research
Technician Berlex Biosciences, South San Francisco; 7/85-2/87.

Worked on a subunit porcine vaccine directed against Mycoplasma hyopneumoniae. Was responsible for generating genomic libraries, screening with degenerate oligonucleotides, and characterizing and expressing clones in E. coli. Also constructed a general purpose expression vector for use by other scientific teams.

PUBLICATIONS

1. Hilary F. Clark, et al. "The Secreted Protein Discovery Initiative (SPDI), a Large-scale Effort to Identify Novel Human Secreted and Transmembrane Proteins: a bioinformatics assessment." Genome Res. Vol 13(10), 2265-2270, 2003

2. Sean H. Adams, Clarissa Chui' Sarah L. Schilbach, Xing Xian Yu, Audrey D. Goddard, J. Christopher Grimaldi, James Lee, Patrick Dowd, David A. Lewin, & Steven Colman' "BFIT, a Unique Acyl-CoA Thioesterase Induced in Thermogenic Brown Adiopose Tissue: Cloning, organization of the humanb gene and assessment of a potential link to obesity" Biochemical Journal, Vol 360, 135-142, 2001

3. Szeto W, Jiang W, Tice DA, Rubinfeld B, Hollingshead PG, Fong SE, Dugger DL, Pham T, Yansura D, Wong TA, Grimaldi JC, Corpuz RT, Singh JS, Frantz GD, , Devaux B, Crowley CW, Schwall RH, Eberhard DA, Rastelli L, Polakis P, and Pennica D. "Overexpression of the Retenoic Acid-Responsive Gene Stra6 in Human Cancers and its Synergistic Activation by Wnt-1 and Retinoic Acid." Cancer Research Vol. 61(10), 4197-4205, 2001

4. Jeanne Kahn, Fuad Mehraban, Gladdys Ingle, Xiaohua Xin, Juliet E. Bryant, Gordon Vehar, Jill Schoenfeld, J. Christopher Grimaldi (incorrectly named as "Grimaldi, CJ"), Franklin Peale, Aparna Draksharapu, David A. Lewin, and Mary E. Gerritsen. "Gene Expression Profiling in an in Vitro Model of Angiogenesis." American Journal of Pathology Vol 156(6), 1887-1900, 2000.

5. Grimaldi JC, Yu NX, Grunig G, Seymour BW, Cottrez F, Robinson DS, Hosken N, Ferlin WG, Wu X, Soto H, O'Garra A, Howard MC, Coffman RL. "Depletion of eosinophils in mice through the use of antibodies specific for C-C chemokine receptor 3 (CCR3). Journal of Leukocyte Biology; Vol. 65(6), 846-53, 1999

6. Oliver AM, Grimaldi JC, Howard MC, Kearney JF. "Independently ligating CD38 and Fc gammaRIIB relays a dominant negative signal to B cells." Hybridoma Vol. 18(2), 113-9, 1999

7. Cockayne DA, Muchamuel T, Grimaldi JC, Muller-Steffner H, Randall TD, Lund FE, Murray R, Schuber F, Howard MC. "Mice deficient for the ecto-nicotinamide adenine dinucleotide glycohydrolase CD38 exhibit altered humoral immune responses." Blood Vol. 92(4), 1324-33, 1998

8. Frances E. Lund, Nanette W. Solvason, Michael P. Cooke, Andrew W. Heath, J. Christopher Grimaldi, Troy D. Randall, R. M. E. Parkhouse, Christopher C Goodnow and Maureen C. Howard. "Signaling through murine CD38 is impaired in antigen receptor unresponsive B cells." European Journal of Immunology, Vol. 25(5), 1338-1345, 1995

9. M. J. Guimaraes, J. F. Bazan, A. Zolotnik, M. V. Wiles, J. C. Grimaldi, F. Lee, T. McClanahan. "A new approach to the study of haematopoietic development in the yolk sac and embryoid body." Development, Vol. 121(10), 3335-3346, 1995

10. J. Christopher Grimaldi, Sriram Balasubramanian, J. Fernando Bazan, Armen Shanafelt, Gerard Zurawski and Maureen Howard. "CD38-mediated protein ribosylation." Journal of Immunology,Vol. 155(2), 811-817, 1995

11. Leopoldo Santos-Argumedo, Frances F. Lund, Andrew W. Heath, Nanette Solvason, Wei Wei Wu, J. Christopher Grimaldi, R. M. E. Parkhouse and Maureen Howard. "CD38 unresponsiveness of xid B cells implicates Bruton's tyrosine kinase (btk) as a regulator of CD38 induced signal transduction." International Immunology, Vol 7(2), 163-170, 1995

12. Frances Lund, Nanette Solvason, J. Christopher Grimaldi, R. M. E. Parkhouse and Maureen Howard. "Murine CD38: An immunoregulatory ectoenzyme." Immunology Today, Vol. 16(10), 469-473, 1995

13. Maureen Howard, J. Christopher Grimaldi, J. Fernando Bazan, Frances E. Lund, Leopoldo Santos-Argumedo, R. M. E. Parkhouse, Timothy F. Walseth, and Hon Cheung Lee. "Formation and Hydrolysis of Cyclic ADP-Ribose Catalyzed by Lymphocyte Antigen CD38." Science, Vol. 262, 1056-1059, 1993

14. Nobuyuki Harada, Leopoldo Santos-Argumedo, Ray Chang, J. Christopher Grimaldi, Frances Lund, Camilynn I. Brannan, Neal G. Copeland, Nancy A. Jenkins, Andrew Heath, R. M. E. Parkhouse and Maureen Howard. "Expression Cloning of a cDNA Encoding a Novel Murine B Cell Activation Marker: Homology to Human CD38." The Journal of Immunology, Vol. 151, 3111-3118, 1993

15. David J. Rawlings, Douglas C. Saffran, Satoshi Tsukada, David A. Largaespada, J. Christopher Grimaldi, Lucie Cohen Randolph N. Mohr, J. Fernando Bazan, Maureen Howard, Neal G. Copeland, Nancy A. Jenkins, Owen Witte. "Mutation of Unique Region of Bruton's Tyrosine Kinase in Immunodeficient XID Mice." Science, Vol. 261, 358-360, 1993

16. J. Christopher Grimaldi, Raul Torres, Christine A. Kozak, Ray Chang, Edward Clark, Maureen Howard, and Debra A. Cockayne. "Genomic Structure and Chromosomal Mapping of the Murine CD40 Gene." The Journal of Immunology, Vol 149, 3921-3926, 1992

17. Timothy C. Meeker, Bruce Shiramizu, Lawrence Kaplan, Brian Herndier, Henry Sanchez, J. Christopher Grimaldi, James Baumgartner, Jacab Rachlin, Ellen Feigal, Mark Rosenblum and Michael S. McGrath. "Evidence for Molecular Subtypes of HIV-Associated Lymphoma:

Division into Peripheral Monoclonal, Polyclonal and Central Nervous System Lymphoma." AIDS, Vol. 5, 669-674, 1991

18. Ann Grimaldi and Chris Grimaldi. "Small-Scale Lambda DNA Prep." Contribution to Current Protocols in Molecular Biology, Supplement 5, Winter 1989

19. J. Christopher Grimaldi, Timothy C. Meeker. "The t(5;14) Chromosomal Translocation in a Case of Acute Lymphocytic Leukemia Joins the Interleukin-3 Gene to the Immunoglobulin Heavy Chain Gene." Blood, Vol. 73, 2081-2085, 1989

20. Timothy C. Meeker, J. Christopher Grimaldi, et al. "An Additional Breakpoint Region in the BCL-1 Locus Associated with the t(11;14) (q13;q32) Translocation of B-Lymphocytic Malignancy." Blood, Vol. 74, 1801-1806, 1989

21 Timothy C. Meeker, J. Christopher Grimaldi, Robert O'Rourke, et al. "Lack of Detectable Somatic Hypermutation in the V Region of the Ig H Chain Gene of a Human Chronic B Lymphocytic Leukemia." The Journal of Immunology, Vol. 141, 3994-3998, 1988

MANUSCRIPTS IN PREPARATION

1. Sriram Balasubramanian, J. Christopher Grimaldi, J. Fernando Bazan, Gerard Zurawski and Maureen Howard. "Structural and functional characterization of CD38: Identification of active site residues"

PATENTS

1. "Methods for Eosinophil Depletion with Antibody to CCR3 Receptor" (US 6,207,155 B1).

2. "Amplification Based Cloning Method." (US 6,607,899)

3. Ashkenazi et al., "Secreted and Transmembrane Polypeptides and Nucleic Acids Encoding the Same." (this patent covers several hundred genes)

4. "IL-17 Homologous Polypeptides and Therapeutic Uses Thereof"

5. "Method of Diagnosing and Treating Cartilaginous Disorders."

MEMBERSHIPS AND ACTIVITIES

Editor	Frontiers in Bioscience
Member	DNAX Safety Committee 1991-1999 Biological Safety Affairs Forum (BSAF) 1990-1991 Environmental Law Foundation (ELF) 1990-1991

The t(5;14) Chromosomal Translocation in a Case of Acute Lymphocytic Leukemia Joins the Interleukin-3 Gene to the Immunoglobulin Heavy Chain Gene

By J. Christopher Grimaldi and Timothy C. Meeker

Chromosomal translocations have proven to be important markers of the genetic abnormalities central to the pathogenesis of cancer. By cloning chromosomal breakpoints one can identify activated proto-oncogenes. We have studied a case of B-lineage acute lymphocytic leukemia (ALL) that was associated with peripheral blood eosinophilia. The chromosomal translocation t(5;14) (q31;q32) from this sample was cloned and studied at the molecular level. This translocation joined the immunoglobulin heavy chain joining (Jh) region to the promotor region of the interleukin-3 (IL-3) gene in opposite transcriptional orientations. The data suggest that activation of the IL-3 gene by the enhancer of the immunoglobulin heavy chain gene may play a central role in the pathogenesis of this leukemia and the associated eosinophilia.

KARYOTYPIC STUDIES of leukemia and lymphoma have identified frequent nonrandom chromosomal translocations. Some of these translocations juxtapose the immunoglobulin heavy chain (IgH) gene with important protooncogenes, such as c-*myc* and *bcl*-2.[1,2] In this way, the IgH gene can activate proto-oncogenes, resulting in disordered gene expression and a step in the development of cancer. The investigation of additional nonrandom translocations into the IgH locus allows us to identify new genes promoting the generation of leukemia and lymphoma.

A distinct subtype of acute lymphocytic leukemia (ALL) has been characterized by B-lineage phenotype, associated eosinophilia in the peripheral blood, and a t(5;14)(q31;q32) chromosomal translocation.[3,4] This syndrome probably occurs in <1% of all patients with ALL. We hypothesized that the cloning of the translocation characteristic of this leukemia might allow the identification of an important gene on chromosome 5 that plays a role in the evolution of this disease. In this report we demonstrate that the interleukin-3 gene (IL-3) and the IgH gene are joined by this translocation.

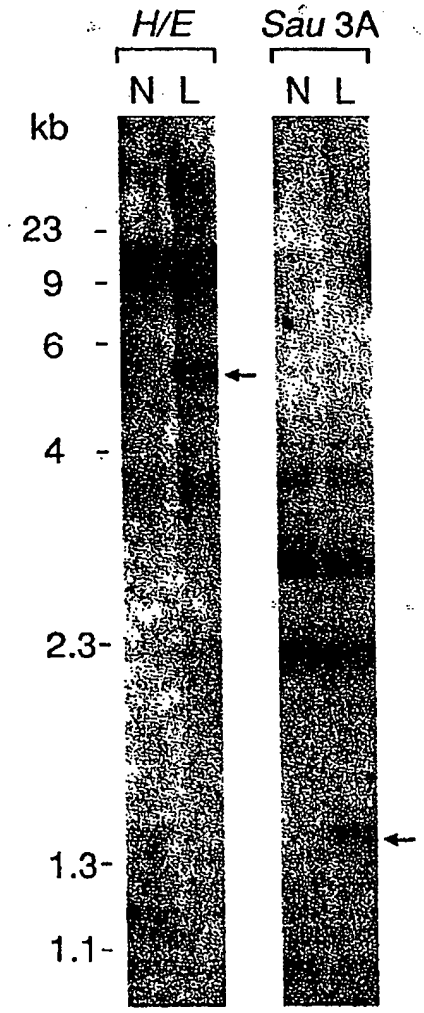


Fig 1. DNA blots of the leukemia sample. The restriction fragment pattern of normal human DNA (N) and the leukemia sample (L) were compared using a human Jh probe. Rearranged bands are indicated by arrows. Sample L exhibits a single rearranged band with both *Hind* III/*Eco*RI and *Sau*3A restriction digests. The rearranged bands are less intense than the other bands because the majority of cells in the sample represent normal bone marrow elements.

MATERIALS AND METHODS

Sample and DNA blots. A bone marrow aspirate from a representative patient with ALL (L1 morphology by French-American-British [FAB] criteria), peripheral eosinophilia (up to 20,000 per microliter with a normal value of <350 per microliter) and a t(5;14)(q31;q32) translocation was studied. Using published methods, genomic DNA was isolated and DNA blots were made.[5] Briefly, 10 μg of high molecular weight (mol wt) DNA were digested using an appropriate restriction enzyme and electrophoresed on a 0.8% agarose gel. The gel was stained with ethidium bromide, photographed, denatured, neutralized, and transferred to Hybond (Amersham, Arlington Heights, IL). After treatment of the filter with ultraviolet light, hybridization was performed. The filter was washed to a final stringency of 0.2% saturated sodium citrate (SSC) and 0.1% sodium lauryl sulfate (SDS) and exposed to film. The human Jh probe has been previously reported.[6]

Genomic library. The genomic library was made using pub-

From the Division of Hematology/Oncology, Department of Medicine, University of California, San Francisco.

Submitted February 22, 1989; accepted March 8, 1989.

Supported by NIH Grant No. CA01102.

Address reprint requests to Timothy C. Meeker, MD, UCSF/VAMC 111H, 4150 Clement St, San Francisco, CA 94121.

Dr Grimaldi's current address is Biostan Inc, 440 Chesapeake Dr, Seaport Centre, Redwood City, CA 94063.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. section 1734 solely to indicate this fact.

0006-4971/89/7308-0031$3.00/0

lished methods.[5] Approximately 100 μg of high mol wt genomic DNA were partially digested with the *Sau*3A restriction enzyme. Fragments from 9 to 23 kilobases (kb) in size were isolated on a sucrose gradient and ligated into phage EMBL3A (Strategene, San Diego). Recombinant phage were packaged, plated, and screened as previously reported.[5]

DNA sequencing. Fragments for sequencing were cloned into M13 vectors and sequenced by the chain termination method using Sequenase (United States Biochemical, Cleveland).[7] All sequence data were derived from both strands.

RESULTS

We studied a bone marrow sample from a patient with ALL and associated peripheral eosinophilia. Karyotypic analysis showed the characteristic t(5;14)(q31;q32) translocation. These features define a distinctive subtype of ALL.[3,4] The leukemic cells were analyzed for cell surface phenotype by immunofluorecence. They were positive for B1 (CD20), B4 (CD19), cALLA (CD10), HLA-DR, and terminal deoxynucleotidyl transferase (Tdt), but negative for surface immunoglobulin. This phenotypic profile describes an immature cell from the B-lymphocytic lineage.[8]

The leukemia DNA was analyzed by Southern blotting for rearrangements of the IgH gene. Using a human immunoglobulin Jh probe, a single rearranged band was detected by *Eco*RI, *Hind*III, *Sst*I, *Sau*3A, and *Eco*RI plus *Hind*III restriction digests, suggesting rearrangement of one allele (Fig 1). The immunoglobulin Jh region from the other allele was presumably either deleted or in the germline configuration.

We hypothesized that the t(5;14)(q31;q32) juxtaposed a growth-promoting gene on chromosome 5 with the immunoglobulin Jh region on chromosome 14. Therefore, a genomic library was made from the leukemic sample and screened with a Jh probe. Fifteen distinct positive clones were isolated and screened for the presence of the rearranged *Sau*3A fragment that was detected by DNA blotting. By this analysis, five clones appeared to represent the rearranged allele identified by DNA blots. One of these clones (clone no. 4) was chosen for further study and a detailed restriction map was generated. The *Eco*RI, *Hind*III/*Eco*RI, and *Sst*I fragments from clone no. 4 that hybridized to the human Jh probe were also identical in size to the rearranged fragments from the leukemia sample, confirming that clone no. 4 represented the rearranged leukemic allele.

Phage clone no. 4 contained 3.7 kb of unknown origin joined to the IgH gene in the region of Jh4 (Fig 2). The IgH gene from Jh4 to the Cmu region appeared to be in germline configuration. Previously, the gene encoding hematopoietic growth factor IL-3 had been mapped to chromosome 5q31 so it was suspected that clone no. 4 might contain part of this gene.[9-12] When the restriction map of human IL-3 and clone no. 4 were compared, they were identical for more than 3 kb (Fig 2).

We confirmed the juxtaposition of the IL-3 gene and the IgH gene by nucleic acid sequencing of the subcloned *Bst*EII/*Hpa*I fragment (Fig 2). The sequence of this fragment showed no disruption of the protein coding region or the messenger RNA of the IL-3 gene. The break in the IL-3 gene occurred in the promotor region, 452 base pairs (bp) upstream of the transcriptional start site (position 64, Fig

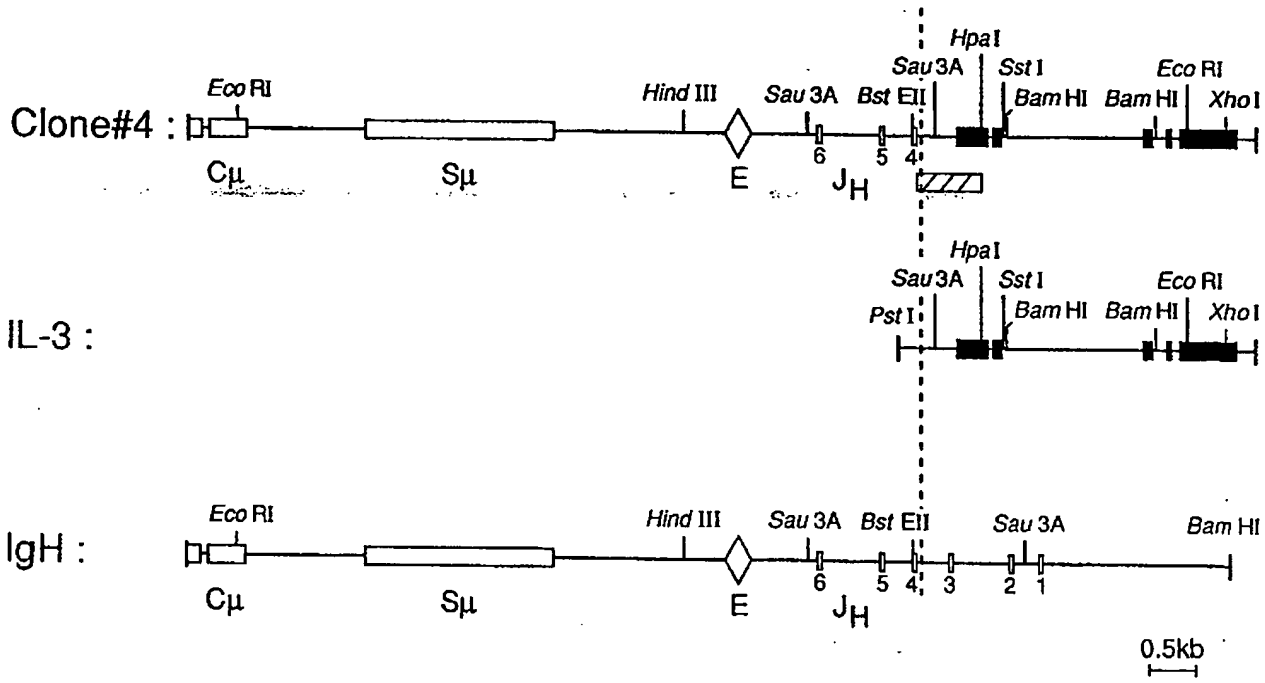


Fig 2. Breakpoint region: t(5;14)(q31;q32). Comparative mapping of phage clone no. 4, the germline IgJh region, and the germline IL-3 gene.[28,29] The map of clone no. 4 is identical to that of IgH until it diverges in the region of Jh4 (at the dashed line), after which it is identical to the map of IL-3. The two genes are positioned in a head-to-head orientation. The Ig μ chain constant region (Cμ), switch region (Sμ), enhancer (E), and Jh segments are indicated (open symbols). The five exons (dark boxes) and four introns of the IL-3 gene are shown. The hatched box indicates the sequenced region.

3A). The break in the IgH gene occurred 2 bp upstream of the Jh4 region. Between the two breaks, 25 bp of uncertain origin (putative N sequence) were inserted.[13,14] No sequences homologous to the immunoglobulin heptamer and nonamer could be identified in the IL-3 sequence (Fig 3B). Therefore, nucleic acid sequencing confirmed the juxtaposition of the IL-3 gene and the IgH gene. The sequence data clearly showed that the genes were positioned in opposite transcriptional orientations (head-to-head).

Available data also allowed us to determine the normal positions of the IL-3 gene and the GM-CSF gene in relation to the centromere of chromosome 5 (Fig 4). The IgH gene is known to be positioned with the variable regions toward the telomere on chromosome 14q.[2,15] It has also been shown that GM-CSF maps within 9 kb of IL-3 in the same transcriptional orientation.[16] Using this information and assuming a simple translocation event in our sample, we can conclude that the IL-3 gene is normally more centromeric, and the GM-CSF gene more telomeric on chromosome 5q (Fig 4). Furthermore, both are transcribed with their 5′ ends toward the centromere.

DISCUSSION

In this report we have cloned a unique chromosomal translocation that appears to be a consistent feature of a rare, yet distinct, clinical form of acute leukemia. This translocation joined the promotor of the IL-3 gene to the IgH gene. Except for the altered promotor, the IL-3 gene appeared

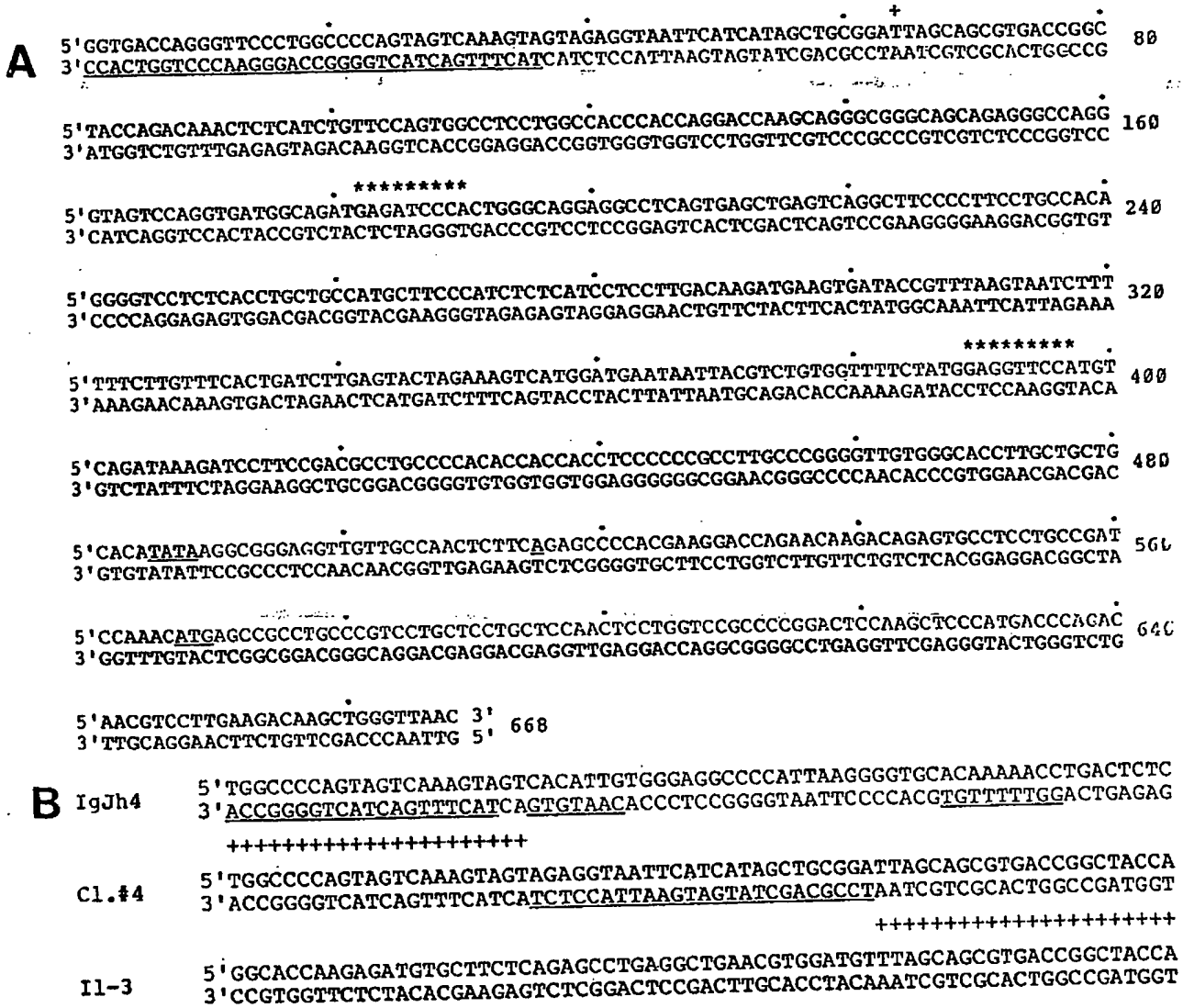


Fig 3. Sequence of t(5;14)(q31;q32) breakpoint region. (A) Nucleotide sequence of the *Bst*EII/*Hpa*I fragment indicated on Fig 2. Nucleotides 1 to 36 represent the Jh4 coding region underlined on the coding strand.[9] Nucleotides 39 to 63 are a putative N region. The sequence from position 64 to 668 is that of the germline IL-3 gene.[26] The IL-3 TATA box (485), transcription start (515), and initiation methionine (567) are underlined. Two proposed regulatory sequences in the promotor are marked by asterisks (positions 182 and 389). (B) Comparative sequence of the t(5;14)(q31;q32) breakpoint region. The IgJh4 region is shown with its coding region, heptamer, and nonamer underlined. Clone no. 4 is shown with putative N region sequences underlined. The IL-3 sequence is also shown. A plus sign (+) denotes the identical nucleotide between sequences. No heptamer or nonamer is identified in the IL-3 sequence.

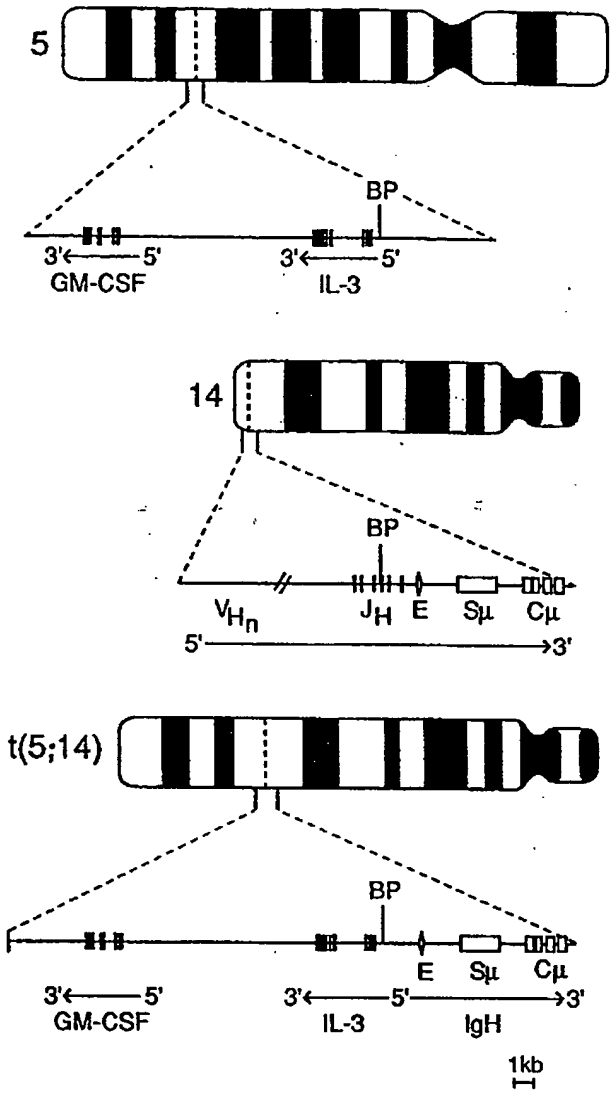


Fig 4. Diagram of the translocation. The normal chromosome 5q31 is shown with the GM-CSF gene telomeric to the IL-3 gene in the transcriptional orientation shown. On normal chromosome 14q32 the Vh regions are telomeric. The t(5;14)(q31;q32) translocation results in the head-to-head orientation of these genes. Symbols are defined in Fig 2. BP, breakpoint position.

intact as no deletions, insertions, or point mutations were detected by restriction mapping of the entire gene and sequencing of part of the gene. The IgH gene has been truncated at the Jh4 region, which places the immunoglobulin enhancer within 2.5 kb of the IL-3 gene.[17,18] This leads to the hypothesis that the enhancer is increasing transcription of a structurally normal IL-3 gene. The same mechanism is important for activation of the c-*myc* gene in some cases of Burkitt's lymphoma.[19] An alternate hypothesis is that the elimination of an upstream IL-3 promotor element is crucial to the activation of the IL-3 gene.

The proposed activation of the IL-3 gene suggests that an autocrine loop is important for the pathogenesis of this leukemia.[20] Over-expression of the IL-3 gene coupled with the presence of the IL-3 receptor in these cells could account for a strong stimulus for proliferation. In this regard, there are data indicating that immature B-lineage lymphocytes and B-lineage leukemias may express the IL-3 receptor.[21,22]

An additional feature of this type of leukemia is the dramatic eosinophilia, consisting of mature forms. It has been hypothesized that the eosinophils do not arise from the malignant clone, but are stimulated by the tumor.[23,24] Because of the known effect of IL-3 on eosinophil differentiation, secretion of high levels of IL-3 by leukemic cells might have a role in the eosinophilia in this type of leukemia.[12]

The data suggest that the recombination mechanism that is active in the IgH gene during normal differentiation has a role in this translocation.[13,14] This is supported by the breakpoint location at the 5′ end of Jh4 and the presence of putative N-region sequences. On the other hand, no recombination signal sequence (heptamer and nonamer) was found in this region on chromosome 5, suggesting that additional factors also played a role. Further studies will elucidate the mechanism of this and other translocations.

In the leukemia we studied, it is possible that the immunoglobulin enhancer also activates the GM-CSF gene, since this gene is probably positioned only 14 kb away (Fig 4). This is known to be within the range of enhancer activation.[25] The interleukin-5 (IL-5) gene maps to chromosome 5q31.[26] Deregulation of the IL-5 gene by this translocation would act synergistically with IL-3 in the stimulation of eosinophil proliferation and differentiation.[27] These and other questions will be answered by the study of more patient samples. We plan to determine whether the t(5;14)(q31;q32) translocation is capable of activating multiple lymphokines simultaneously and whether they cooperate in the generation of this leukemia.

REFERENCES

1. Klein G, Klein E: Evolution of tumours and the impact of molecular oncology. Nature 315:190, 1985

2. Showe L, Croce C: The role of chromosomal translocations in B- and T-cell neoplasia. Annu Rev Immunol 5:253, 1987

3. Hogan T, Koss W, Murgo A, Amato R, Fontana J, VanScoy F: Acute lymphoblastic leukemia with chromosomal 5;14 translocation and hypereosinophilia: Case report and literature review. J Clin Oncol 5:382, 1987

4. Tono-oka T, Sato Y, Matsumoto T, Ueno N, Ohkawa M, Shikano T, Takeda T: Hypereosinophilic syndrome in acute lymphoblastic leukemia with a chromosome translocation t(5q;14q). Med Pediatr Oncol 12:33, 1984

5. Meeker T, Grimaldi JC, O'Rourke R, Loeb J, Juliusson G, Einhorn S: Lack of detectable somatic hypermutation in the V region of the IgH gene of a human chronic B-lymphocytic leukemia. J Immunol 141:3394, 1988

6. Ravetch J, Siebenlist U, Korsmeyer S, Waldmann T, Leder P: Structure of the human immunoglobulin μ locus: Characterization of embryonic and rearranged J and D genes. Cell 27:583, 1981

7. Norrander U, Kempe T, Messing J, Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene 26:101, 1983

8. Foon K, Todd R: Immunologic classification of leukemia and lymphoma. Blood 68:1, 1986

9. LeBeau M, Epstein N, O'Brien SJ, Nienhuis AW, Yang Y-C, Clark S, Rowley J: The interleukin-3 gene is located on human

chromosome 5 and is deleted in myeloid leukemias with a deletion of 5q. Proc Natl Acad Sci USA 84:5913, 1987

10. LeBeau M, Chandrasekharappi S, Lemons R, Schwartz J, Larson R, Arai N, Westbrook C: Molecular and cytogenetic analysis of chromosome 5 abnormalities in myeloid disorders, in cancer cells, in Proceedings of Molecular Diagnostics of Human Cancer. Cold Spring Harbor Laboratory, NY, 1989 (in press)

11. Ihle J, Weinstein Y: Immunological regulation of hematopoietic/lymphoid stem cell differentiation by interleukin-3. Adv Immunol 39:1, 1986

12. Clark S, Kamen R: The human hematopoietic colony-stimulating factors. Science 236:1229, 1987

13. Bakhshi A, Wright J, Graninger W, Seto M, Owens J, Cossman J, Jensen J, Goldman P, Korsmeyer S: Mechanism of the t(14,18) chromosomal translocation: Structural analysis of both derivative 14 and 18 reciprocal partners. Proc Natl Acad Sci USA 84:2396, 1987

14. Tsujimoto Y, Louie E, Bashir M, Croce C: The reciprocal partners of both the t(14,18) and the t(11,14) translocations involved in B-cell neoplasms are rearranged by the same mechanism. Oncogene 2:347, 1988

15. Erikson J, Finan J, Nowell P, Croce C: Translocation of immunoglobulin VH genes in Burkitt lymphoma. Proc Natl Acad Sci USA 80:810, 1982

16. Yang Y-C, Kovacic S, Kriz R, Wolf S, Clark S, Wellems T, Nienhuis A, Epstein N: The human genes for GM-CSF and IL-3 are closely linked in tandem on chromosome 5. Blood 71:958, 1988

17. Gillies S, Morrison S, Oi V, Tonegawa S: A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell 33:717, 1983

18. Banerji J, Olson L, Schaffner W: A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell 33:729, 1983

19. Hayday A, Gillies S, Saito H, Wood C, Wiman K, Hayward W, Tonegawa S: Activation of a translocated human c-*myc* gene by an enhancer in the immunoglobulin heavy-chain locus. Nature 307:334, 1984

20. Sporn M, Roberts A: Autocrine growth factors and cancer. Nature 313:745, 1985

21. Palacios R, Steinmetz M: IL-3–dependent mouse clones that express B-220 surface antigen, contain Ig genes in germ line configuration, and generate B lymphocytes in vivo. Cell 41:727, 1985

22. Uckun F, Gesner T, Song C, Myers D, Mufson A: Leukemic B-cell precursors express functional receptors for human interleukin-3. Blood 73:533, 1989

23. Spitzer G, Garson O: Lymphoblastic leukemia with marked eosinophilia: A report of two cases. Blood 42:377, 1973

24. Catovsky D, Bernasconi C, Verkonck P, Postma A, Howss J, Berg A, Rees J, Castelli G, Morra E, Galton D: The association of eosinophilia with lymphoblastic leukemia or lymphoma: A study of seven patients. Br J Haematol 45:523, 1980

25. Wang X-F, Calame K: The endogenous immunoglobulin heavy chain enhancer can activate tandem Vh promoters separated by a large distance. Cell 43:659, 1985

26. Sutherland G, Baker E, Callen D, Campbell H, Young I, Sanderson C, Garson O, Lopez A, Vadas M: Interleukin-5 is at 5q31 and is deleted in the 5q-syndrome. Blood 71:1150, 1988

27. Warren D, Moore M: Synergism among interleukin-1, interleukin-3, and interleukin-5 in the production of eosinophils from primitive hemopoietic stem cells. J Immunol 140:94, 1988

28. Yang Y-C, Clark S: Molecular cloning of a primate cDNA and the human gene for interleukin-3. Lymphokines 15:375, 1988

29. Yang Y-C, Ciarletta A, Temple P, Chung M, Kovacic S, Witek-Giannotti J, Leary A, Kriz R, Donahue R, Wong G, Clark S: Human IL-3 (multi-CSF): Identification by expression cloning of a novel hematopoietic growth factor related to murine IL-3. Cell 47:3, 1986

RAPID COMMUNICATION

Activation of the Interleukin-3 Gene by Chromosome Translocation in Acute Lymphocytic Leukemia With Eosinophilia

By Timothy C. Meeker, Dan Hardy, Cheryl Willman, Thomas Hogan, and John Abrams

The t(5;14)(q31;q32) translocation from B-lineage acute lymphocytic leukemia with eosinophilia has been cloned from two leukemia samples. In both cases, this translocation joined the IgH gene and the interleukin-3 (IL-3) gene. In one patient, excess IL-3 mRNA was produced by the leukemic cells. In the second patient, serum IL-3 levels were measured and shown to correlate with disease activity. There was no evidence of excess granulocyte/macrophage colony stimulating factor (GM-CSF) or IL-5 expression. Our data support the formulation that this subtype of leukemia may arise in part because of a chromosome translocation that activates the IL-3 gene, resulting in autocrine and paracrine growth effects.

A NUMBER OF chromosome translocations have been associated with human leukemia and lymphoma. In many cases the study of these translocations has led to the discovery or characterization of proto-oncogenes, such as *bcl*-2, c-*abl*, and c-*myc*, that are located adjacent to the translocation.[1,2] It is now widely understood that cancer-associated translocations disrupt nearby proto-oncogenes.

A distinct subtype of acute leukemia is characterized by the triad of B-lineage immunophenotype, eosinophilia, and the t(5;14)(q31;q32) translocation.[3,4] Leukemic cells from such patients have been positive for terminal deoxynucleotidyl transferase (Tdt), common acute lymphoblastic leukemia antigen (CALLA), and CD19, but negative for surface or cytoplasmic immunoglobulin. In previous work, we cloned the t(5;14) breakpoint from one leukemic sample (Case 1) and determined that the IgH and interleukin-3 (IL-3) genes were joined by this abnormality.[5] In this report, we extend those findings by showing that the t(5;14)(q31;q32) translocation from a second leukemia sample (Case 2) has a similar structure, and we report our study of growth factor expression in these patients.

MATERIALS AND METHODS

Samples and Southern blots. Case 1 has been described.[5,6] Clinical features of Case 2 have been described in detail.[3] DNA isolation and Southern blotting was done using previously described methods.[5] Filters were hybridized with an immunoglobulin Jh probe, a 280 bp *Bam*HI/*Eco*RI genomic IL-3 fragment, and an IL-3 cDNA probe.[7,8]

Northern blots. RNA isolation and Northern blotting have been described.[9] Briefly, Northern blots were done by separating 9μg total RNA on 1% agarose-formaldehyde gels. Equal RNA loading in each lane was confirmed by ethidium bromide staining. Blots were hybridized with an IL-3 cDNA probe extending to the *Xho* I site in exon 5, a 720 bp *Sst* I/*Kpn* I probe derived from intron 2 of the IL-3 gene, a 600 bp *Nhe* I/*Hpa* I IL-5 cDNA probe, and a 500 bp *Pst* I/*Nco* I granulocyte-macrophage colony stimulating factor (GM-CSF) cDNA probe.[10-12]

Polymerase chain reaction. Primers were designed with *Bam*HI sites for cloning. One primer hybridized to the Jh sequences from the IgH gene (Primer 144:5'-TAGGATCCGACGGTGACCAGGGT), and the other hybridized to the region of the TATA box in the IL-3 gene (Primer 161: 5'-AACAGGATCCCGCCTTATATGTGCAG). Polymerase chain reaction (PCR) (95°C for 1 minute, 61°C for 30 seconds, and 72°C for 3 minutes) was done using 500 ng genomic DNA and 50 pmol of each primer in 100 μL containing 67 mmol/L Tris-HCl pH 8.8, 6.7 mmol/L $MgCl_2$, 10% dimethyl sulfoxide (DMSO), 170 μg/mL bovine serum albumin (BSA) (fraction V), 16.6 mmol/L ammonium sulfate, 1.5 mmol/L each dNTP and Taq polymerase (Perkin-Elmer, Norwalk, CT).[13]

Sequencing. Sequencing was done by chain termination in M13 vectors.[14] As part of this study, we sequenced a subclone of a normal IL-3 promotor, covering 598 base pairs from a *Sma* I site at position −1240 (with respect to the proposed site of transcription initiation) to an *Nhe* I site at position −642. The plasmid containing this region was a gift from Naoko Arai of the DNAX Research Institute.

Expression in Cos7 cells. A genomic IL-3 fragment from Case 1 was cloned into the pXM expression vector.[10] Briefly, the *Hin*dIII/*Sal* I fragment containing the IL-3 gene was subcloned from the previously described phage clone 4 into pUC18.[5] The 2.6 kb fragment extending from the *Sma* I site 61 bp upstream of the IL-3 transcription start to the *Sma* I site in the polylinker was cloned into the blunted *Xho* I site of pXM. The negative control construct was the pXM vector without insert. Plasmids were introduced into Cos7 cells by electroporation, and supernatant was collected after 48 hours in culture.

TF1 bioassay. TF-1 cells were passaged in RPMI 1640 supplemented with 10% heat-inactivated fetal bovine serum, 2 mmol L-glutamine, and 1 ng/mL human GM-CSF.[15] Samples and antibodies were diluted in this same medium lacking GM-CSF but containing penicillin and streptomycin. A 25 μL volume of serial dilutions of patient serum was added to wells in a flat bottom 96-well microtiter plate. Rat anti-cytokine monoclonal antibody in a volume of 25 μL was added to appropriate wells and preincubated for 1 hour at 37°C. Fifty microliters of twice washed TF-1 cells were added to each well, giving a final cell concentration of 1×10^4 cells per well (final volume, 100 μL). The plate was incubated for 48 hours. The remaining cell viability was determined metabolically by the colori-

From the Division of Hematology/Oncology 111H, Department of Medicine, University of California and the Veterans Administration Medical Center, San Francisco, CA; the Center for Molecular and Cellular Diagnostics, Department of Pathology and Cell Biology, University of New Mexico, Albuquerque, NM; the Division of Hematology/Oncology, Department of Medicine, West Virginia University, Morgantown, WV; and DNAX Research Institute, Palo Alto, CA.

Submitted March 27, 1990; accepted April 19, 1990.

Supported in part by the University of California Cancer Research Coordinating Committee and University of New Mexico Cancer Center funding from the state of New Mexico. The DNAX Research Institute is supported by Schering-Plough.

Address reprint requests to Timothy C. Meeker, MD, Division of Hematology/Oncology 111H, Department of Medicine, University of California and the Veterans Administration Medical Center, 4150 Clement St, San Francisco, CA 94121.

0006-4971/90/7602-0022$3.00/0

```
IgJh5   5'GACGGTGACCAGGGTTCCTTGGCCCCAGGAGTCGAACCAGTTGTCACATTGTGACAACAATGCCAGACCCCGACAAAGAACGGGGCCCCCGCACGTCCCT
        3'CTGCCACTGGTCCCAAGGAACCGGGGTCCTCAGCTTGGTCAACAGTGTAACACTGTTGTTACGGTCTGGGGCTGTTTCTTGCCCGCGGGGCGTGCAGGGA
          ++++++++++++++++++ ++++++++++ +++++++++++++++

Clone   5'GACGGTGACCAGGGTTCCCTGGCCCCAGGGGTCGAACCAGTTGTACGGCCATGTTATTGGGGGGGATCAAGACCTCTCAATACAACTGTCTCCTGCCAAT
        3'CTGCCACTGGTCCCAAGGGACCGGGGTCCCCAGCTTGGTCAACATGGCGGTACAATAACCCCCCCTAGTTCTGGAGAGTTATGTTGACAGAGGACGGTTA
                                                                    (-934)*++++++++++++++++++++++++++++++++++++++++

IL3     5'CCCCTCTCTGCAAACCTTGCCTACTGGGCCTGCACCTGGCAAATCCATGCTCAGCACAGACGGGGATCAAGACCTCTCAATACAACTGTCTCCTGCCAAT
        3'GGGGAGAGACGTTTGGAACGGATGACCCGGACGTGGACCGTTTAGGTACGAGTCGTGTCTGCCCCTAGTTCTGGAGAGTTATGTTGACAGAGGACGGTTA
```

Fig 1. Breakpoint sequences for Case 2. The germline IgJh5 region sequence (protein coding region and recombination signal sequences are underlined) is on top, the translocation sequence from Case 2 (PCR primer sequences and putative N region are underlined) is in the middle, and the germline IL-3 sequence, which we derived from a normal IL-3 clone, is on the bottom.[7] + indicates that each sequence has the same nucleotide. The sequence documents the head-to-head joining of the IL-3 and IgH genes. The breakpoint in the IL-3 gene occurred at position −934 (*).

metric method of Mosmann using a VMax microtiter plate reader (Molecular Devices, Menlo Park, CA) set at 570 and 650 nm.[16]

Cytokine immunoassays. These assays used rat monoclonal anti-cytokine antibodies (10 μg/mL) to coat the wells of a PVC microtiter plate. The capture antibodies used were BVD3-6G8, JES1-39D10, and BVD2-23B6, for the IL-3, IL-5, and GM-CSF assays, respectively. Patient sera were then added (undiluted and diluted 1:2 for IL-3, undiluted for IL-5, and undiluted and diluted 1:5 for GM-CSF). The detecting immunoreagents used were either mouse antiserum to IL-3 or nitroiodophenyl (NIP)-derivatized rat monoclonal antibodies JES1-5A2 and BVD2-21C11, specific for IL-5 and GM-CSF, respectively. Bound antibody was subsequently detected with immunoperoxidase conjugates: horseradish peroxidase (HRP)-labeled goat anti-mouse Ig for IL-3, or HRP-labeled rat (J4 MoAb) anti-NIP for IL-5 and GM-CSF. The chromogenic substrate was 3-3'azino-bis-benzthiazoline sulfonate (ABTS; Sigma, St Louis, MO). Unknown values were interpolated from standard curves prepared from dilutions of the recombinant factors using Softmax software available with the VMAX microplate reader (Molecular Devices).

RESULTS

Leukemic DNA from Case 2 was studied by Southern blotting. When digested with the *Hin*dIII restriction enzyme and hybridized with a human immunoglobulin heavy chain joining region (Jh) probe, a rearranged fragment at approximately 14 kb was detected (data not shown). When reprobed with either of two different IL-3 probes, a rearranged 14 kb fragment, comigrating with the rearranged Jh fragment, was identified. When leukemic DNA was digested with *Hin*dIII plus *Eco*RI, a rearranged Jh fragment was detected at 6 kb. The IL-3 probes also identified a comigrating fragment of this size. These experiments indicated that the leukemic sample studied was clonal and that a single fragment contained both Jh and IL-3 sequences, suggesting a translocation had occurred.

To characterize better the joining of the IL-3 gene and the immunoglobulin heavy chain (IgH) gene, the polymerase chain reaction (PCR) was used to clone the translocation.[13] A Jh primer and an IL-3 primer were designed to produce an amplified product in the event of a head-to-head translocation. While control DNA gave no PCR product, Case 2 DNA yielded a PCR-derived fragment of approximately 980 bp, which was cloned and sequenced.

The DNA sequence of the translocation clone from Case 2 confirmed the joining of the Jh region with the promotor of the IL-3 gene in a head-to-head configuration (Fig 1). Sequence analysis indicated that the breakpoint on chromosome 14 was just upstream of the Jh5 coding region. The breakpoint on chromosome 5 occurred 934 bp upstream of the putative site of transcription initiation of the IL-3 gene. We also determined that a putative N sequence of 17 bp was inserted between the chromosome 5 and chromosome 14 sequences during the translocation event.[17,18] Figure 2 shows

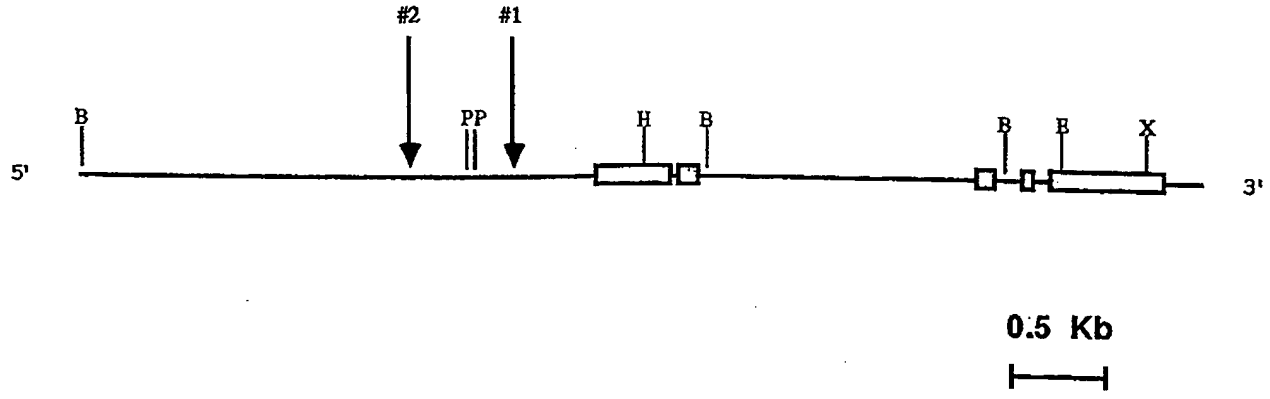


Fig 2. Relationship of chromosome 5 breakpoints to the IL-3 gene. This figure shows the two cloned breakpoints (arrows) in relation to the normal IL-3 gene.[8,10] One breakpoint occurred at position −462 and the other at −934 (arrows). In both circumstances, the translocations resulted in a head-to-head joining of the IgH gene and the IL-3 gene, leaving the mRNA and protein coding regions of the IL-3 gene intact. Boxes denote the five IL-3 exons; restriction enzymes are (B) *Bam*HI, (P) *Pst* I, (H) *Hpa* I, (E) *Eco*RI, and (X) *Xho* I.

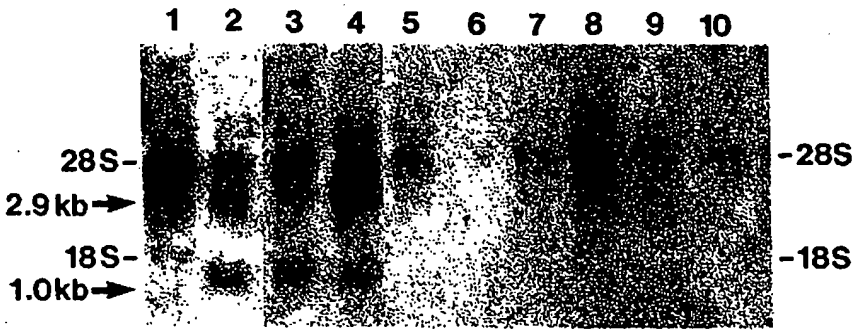


Fig 3. Documentation of IL-3 mRNA over-expression. A Northern blot was prepared and hybridized with a probe for IL-3. Lane 1 contained RNA from unstimulated peripheral blood lymphocytes (PBL) as a negative control. Lane 2 contained RNA from PBL stimulated for 4 hours with concanavalian A (ConA), and lane 3 contained RNA from PBL stimulated with ConA for 48 hours. As in the positive control lanes (2 and 3), a 1 kb band was identified in the leukemic sample from Case 1 (lane 4, lower arrow), suggesting aberrant expression of the IL-3 gene. In addition, the leukemic sample showed over-expression of an unspliced 2.9 kb IL-3 transcript (lane 4, upper arrow). We documented that this represented an unspliced precurser of the mature 1 kb transcript by showing that this band hybridized to a probe from intron 2 of the IL-3 gene. A similar 2.9 kb band was detected in lane 2, suggesting that an IL-3 mRNA of this size is sometimes detectable in normal mitogen-stimulated cells. Lane 5 through 10 represent RNA from six samples of B-lineage acute lymphocytic leukemia without the t(5;14) translocation, indicating that only the sample with the translocation exhibited IL-3 over-expression. Case 2 could not be analyzed by Northern blot because too few cells were available for study.

the locations of the two cloned breakpoints in relation to the IL-3 gene. The two chromosome 5 breakpoints were separated by less than 500 bp.

The genomic structure in Cases 1 and 2 suggested that a normal IL-3 gene product was over-expressed as a result of the altered promotor structure. This would predict that the IL-3 gene on the translocated chromosome was capable of making IL-3 protein. This prediction was tested by expressing a genomic fragment from the translocated allele of Case 1 containing all five IL-3 exons under the control of the SV40 promotor/enhancer in the Cos7 cell line. Cell supernatants were studied in a proliferation assay using the factor dependent erythroleukemic cell line, TF-1. The supernatants derived from transfections using the vector plus insert supported TF-1 proliferation, while supernatants from transfections using the vector alone were negative in this assay (data not shown). Furthermore, the biologic activity could be blocked by an antibody to human IL-3 (BVD3-6G8). This result showed that the translocated allele retained the ability to make IL-3 mRNA and protein.

The level of expression of IL-3 mRNA in leukemic cells from Case 1 was assessed. Northern blotting showed that the mature IL-3 mRNA (approximately 1 kb) and a 2.9 kb unspliced IL-3 mRNA were excessively produced by the leukemia (Fig 3). The 2.9 kb form of the mRNA is also present at low levels in normal peripheral blood T lymphocytes after mitogen activation (Fig 3). Several B-lineage acute leukemia samples without the t(5;14) translocation had undetectable levels of IL-3 mRNA in these experiments. In addition, although genes for GM-CSF and IL-5 map close to the IL-3 gene and might have been deregulated by the translocation, no IL-5 or GM-CSF mRNA could be detected in the leukemic sample (data not shown).[19,20]

Three serum samples from Case 2 were assayed by immunoassay for levels of IL-3, GM-CSF, and IL-5 (Table 1). Serum IL-3 could be detected and correlated with the clinical course. When the patient's leukemic cell burden was highest, the IL-3 level was highest. No serum GM-CSF or IL-5 could be detected.

Since the IL-3 immunoassay measured only immunoreactive factor, we confimed that biologically active IL-3 was present by using the TF-1 bioassay. This bioassay can be rendered monospecific using appropriate neutralizing monoclonal antibodies specific for IL-3, IL-5, or GM-CSF. We observed that sera from 1-16-84 and 3-14-84 contained TF-1 stimulating activity that could be blocked with anti–IL-3 MoAb (BVD3-6G8), but not with MoAbs to IL-5 (JES1-39D10) or GM-CSF (BVD2-23B6) (Fig 4; GM-CSF data not shown). The amount of neutralizable bioactivity in these two samples correlated very well with the difference in IL-3 levels obtained by immunoassay for these samples. Furthermore, the failure to block TF-1 proliferating activity with either anti–IL-5 or anti–GM-CSF was consistent with the inability to measure these factors by immunoassay and

Table 1. Peripheral Blood Counts and Growth Factor Levels at Different Times in Case 2

	Sample Date		
	11/15/83	1/16/84	3/14/84
Peripheral blood counts (cells/μL)			
WBC	81,800	116,500	12,300
Lymphoblasts	0	33,785	0
Eosinophils	46,626	73,080	615
Serum growth factor levels (pg/mL)			
IL-3	<444	7,995	1,051
GM-CSF	<15	<15	<15
IL-5	<50	<50	<50

Peripheral blood counts from Case 2 at three different time points with the corresponding growth factor levels quantified by immunoassay. The patient received chemotherapy between 1/16/84 and 3/14/84 to lower his leukemic burden.[3] No serum samples were available for a similar analysis of Case 1.

Abbreviation: WBC, white blood cells.

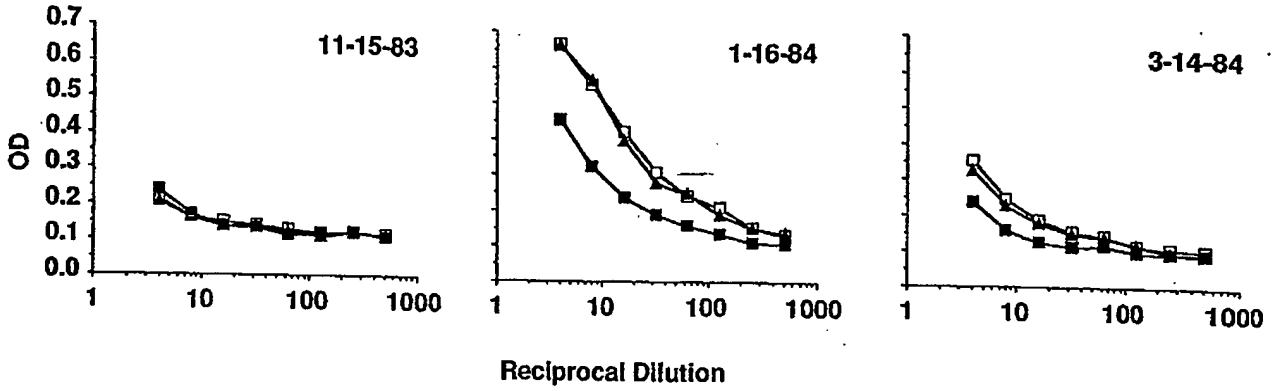


Fig 4. Bioassay of serum IL-3. Leukemic patient sera were tested for bioactive IL-3 and IL-5 in the TF-1 proliferation assay. The reciprocal of the dilution is indicated on the horizontal axis and the optical density indicating the amount of proliferation is indicated on the vertical axis. Serum from all three time points was assayed simultaneously. The assay was rendered monospecific by using a 1 μg/mL final concentration of monoclonal rat anti–IL-3, BVD3-6G8 (■), or anti–IL-5, JES1-39D10 (▲); □ indicates no MoAb. On 1/16/84 and 3/14/84, inhibition of proliferation was evident in the presence of anti–IL-3 antibody, documenting serum levels of IL-3 on those days. Serum IL-5 was not detected in this assay, as anti–IL-5 did not alter TF-1 proliferation.

indicated that these other myeloid growth factors were not detectably circulating in the serum of this patient.

DISCUSSION

In this report, we have extended our analysis of acute lymphocytic leukemia and eosinophilia associated with the t(5;14) translocation. In both cases we have studied, we have documented the joining of the IL-3 gene from chromosome 5 to the IgH gene from chromosome 14. The breakpoints on chromosome 5 are within 500 bp of each other, suggesting that additional breakpoints will be clustered in a small region of the IL-3 promotor. The PCR assay we have developed will be useful in the screening of additional clinical samples for this abnormality.

The finding of a disrupted IL-3 promotor associated with an otherwise normal IL-3 gene implied that this translocation might lead to the over-expression of a normal IL-3 gene product. In this work, we have documented that this is true. In addition, neither GM-CSF nor IL-5 are over-expressed by the leukemic cells. Furthermore, in one patient, serum IL-3 could be measured and correlated with disease activity. To our knowledge, this is the first measurement of human IL-3 in serum and its association with a disease process. The measurement of serum IL-3 in this and other clinical settings may now be indicated.

The finding of the IL-3 gene adjacent to a cancer-associated translocation breakpoint suggests that its activation is important for oncogenesis. It is our thesis that an autocrine loop for IL-3 is important for the evolution of this leukemia.[21] The excessive IL-3 production that we have documented would be one feature of such an autocrine loop. The final proof of our thesis must await additional data. In particular, from the study of additional clinical samples, it will be necessary to document that the IL-3 receptor is present on the leukemic cells and that anti–IL-3 antibody decreases proliferation of the leukemia in vitro.

An important aspect of this work is the suggestion of a therapeutic approach for this disease. If an autocrine loop for IL-3 can be documented in this disease, attempts to lower circulating IL-3 levels or block the interaction of IL-3 with its receptor may prove useful. Because it is also possible that the eosinophilia in these patients is mediated by the paracrine effects of leukemia-derived IL-3, similar interventions may improve this aspect of the disease. Antibodies or engineered ligands to accomplish these goals may soon be available.

ACKNOWLEDGMENT

We thank Naoko Arai, Ken-ichi Arai, R. O'Rourke, J. Grimaldi, and T. O'Connell for technical assistance and/or helpful discussions.

REFERENCES

1. Klein G, Klein E: Evolution of tumours and the impact of molecular oncology. Nature 315:190, 1985

2. Showe L, Croce C: The role of chromosomal translocations in B- and T-cell neoplasia. Ann Rev Immunol 5:253, 1987

3. Hogan T, Koss W, Murgo A, Amato R, Fontana J, VanScoy F: Acute lymphoblastic leukemia with chromosomal 5;14 translocation and hypereosinophilia: Case report and literature review. J Clin Oncol 5:382, 1987

4. Tono-oka T, Sato Y, Matsumoto T, Ueno N, Ohkawa M, Shikano T, Takeda T: Hypereosinophilic syndrome in acute lymphoblastic leukemia with a chromosome translocation t(5q;14q). Med Ped Oncol 12:33, 1984

5. Grimaldi J, Meeker T: The t(5;14) chromosomal translocation in a case of acute lymphocytic leukemia joins the interleukin-3 gene to the immunoglobulin heavy chain gene. Blood 73:2081, 1989

6. McConnell T, Foucar K, Hardy W, Saiki J: Three-way reciprocal chromosomal translocation in an acute lymphoblastic leukemia patient with hypereosinophilia syndrome. J Clin Oncol 5:2042, 1987

7. Ravetch J, Siebenlist U, Korsmeyer S, Waldmann T, Leder P: Structure of the human immunoglobulin m locus: Characterization of embryonic and rearranged J and D genes. Cell 27:583, 1981

8. Otsuka T, Miyajima A, Brown N, Otsu K, Abrams J, Saeland S, Caux C, Malefijt R, Vries J, Meyerson P, Yokota K, Gemmel L,

Rennick D, Lee F, Arai N, Arai K, Yokota T: Isolation and characterization of an expressible cDNA encoding human IL-3. J Immunol 140:2288, 1988

9. Sambrook J, Fritsch E, Maniatis T: Molecular Cloning. Cold Spring Harbor, NY, Cold Spring Harbor Press, 1989

10. Yang Y-C, Ciarletta A, Temple P, Chung M, Kovacic S, Witek-Giannotti J, Leary A, Kriz R, Donahue R, Wong G, Clark S: Human IL-3 (multi-CSF): Identification by expression cloning of a novel hematopoietic growth factor related to murine IL-3. Cell 47:3, 1986

11. Yokota T, Coffman R, Hagiwara H, Rennick D, Takebe Y, Yokota K, Gemmell L, Shrader B, Yang G, Meyerson P, Luh J, Hoy P, Pene J, Briere F, Spits H, Banchereau J, Vries J, Lee F, Arai N, Arai K: Isolation and characterization of lymphokine cDNA clones encoding mouse and human IgA-enhancing factor and eosinophil colony-stimulating factor activities: Relationship to interleukin 5. Proc Natl Acad Sci USA 84:7388, 1987

12. Wong G, Witek J, Temple P, Wilkens K, Leary A, Luxenberg D, Jones S, Brown E, Kay R, Orr E, Shoemaker C, Golde D, Kaufman R, Hewick R, Wang E, Clark S: Human GM-CSF: Molecular cloning of the complementary DNA and purification of the natural and recombinant proteins. Science 228:810, 1985

13. Saiki R, Scharf S, Faloona F, Mullis K, Horn G, Erlich H, Arnheim N: Enzymatic amplification of B-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230:1350, 1985

14. Norrander U, Kempe T, Messing J: Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene 26:101, 1983

15. Kitamura T, Tange T, Terasawa T, Chiba S, Kuwaki T, Miyagawa K, Piao Y, Miyazono K, Urabe A, Takaku F: Establishment and characterization of a unique human cell line that proliferates dependently on GM-CSF, IL-3, or erythropoietin. J Cell Physiol 140:323, 1989

16. Mosmann T: Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. J Immunol Methods 65:55, 1983

17. Bakhshi A, Wright J, Graninger W, Seto M, Owens J, Cossman J, Jensen J, Goldman P, Korsmeyer S: Mechanism of the t(14;18) chromosomal translocation: Structural analysis of both derivative 14 and 18 reciprocal partners. Proc Natl Acad Sci USA 84:2396, 1987

18. Tsujimoto Y, Louie E, Bashir M, Croce C: The reciprocal partners of both the t(14;18) and the t(11;14) translocations involved in B-cell neoplasms are rearranged by the same mechanism. Oncogene 2:347, 1988

19. Yang Y-C, Kovacic S, Kriz R, Wolf S, Clark S, Wellems T, Nienhuis A, Epstein N: The human genes for GM-CSF and IL-3 are closely linked in tandem on chromosome 5. Blood 71:958, 1988

20. Sutherland G, Baker E, Callen D, Campbell H, Young I, Sanderson C, Garson O, Lopez A, Vadas M: Interleukin-5 is at 5q31 and is deleted in the 5q− syndrome. Blood 71:1150, 1988

21. Sporn M, Roberts A: Autocrine growth factors and cancer. Nature 313:745, 1985

Clinical and Pathologic Significance of the c-*erb*B-2 (*HER*-2/*neu*) Oncogene

Timothy P. Singleton and John G. Strickler

The c-*erb*B-2 oncogene was first shown to have clinical significance in 1987 by Slamon et al,[79] who reported that c-*erb*B-2 DNA amplification in breast carcinomas correlated with decreased survival in patients with metastasis to axillary lymph nodes. Subsequent studies, however, of c-*erb*B-2 activation in breast carcinoma reached conflicting conclusions about its clinical significance. This oncogene also has been reported to have clinical and pathologic implications in other neoplasms. Our review summarizes these various studies and examines the clinical relevance of c-*erb*B-2 activation, which has not been emphasized in recent reviews.[37,38,55] The molecular biology of the c-*erb*B-2 oncogene has been extensively reviewed[37,39,55] and will be discussed only briefly here.

BACKGROUND

The c-*erb*B-2 oncogene was discovered in the 1980s by three lines of investigation. The *neu* oncogene was detected as a mutated transforming gene in neuroblastomas induced by ethylnitrosurea treatment of fetal rats.[6,73,74,78] The c-*erb*B-2 was a human gene discovered by its homology to the retroviral gene v-*erb*B.[33,49,76] *HER*-2 was isolated by screening a human genomic DNA library for homology with v-*erb*B.[24] When the DNA sequences were determined subsequently, c-*erb*B-2, *HER*-2, and *neu* were found to represent the same gene. Recently, the c-*erb*B-2 oncogene also has been referred to as *NGL*.

The c-*erb*B-2 DNA is located on human chromosome 17q21[24,33,66] and codes for c-*erb*B-2 mRNA (4.6 kb), which translates c-*erb*B-2 protein (p185). This

protein is a normal component of cytoplasmic membranes. The c-*erb*B-2 oncogene is homologous with, but not identical to, c-*erb*B-1, which is located on chromosome 7 and codes for the epidermal growth factor receptor.[8,103] The c-*erb*B-2 protein is a receptor on cell membranes and has intracellular tyrosine kinase activity and an extracellular binding domain.[2,105] Electron microscopy with a polyclonal antibody detects c-*erb*B-2 immunoreactivity on cytoplasmic membranes of neoplasms, especially on microvilli and the non-villous outer cell membrane.[61] In normal cells, immunohistochemical reactivity for c-*erb*B-2 is frequently present at the basolateral membrane or the cytoplasmic membrane's brush border.[22,62]

There is experimental evidence that c-*erb*B-2 protein may be involved in the pathogenesis of breast neoplasia. Overproduction of otherwise normal c-*erb*B-2 protein can transform a cell line into a malignant phenotype.[25] Also, when the *neu* oncogene containing an activating point mutation is placed in transgenic mice with a strong promoter for increased expression, the mice develop multiple independent mammary adenocarcinomas.[18,63] In other experiments, monoclonal antibodies against the *neu* protein inhibit the growth (in nude mice) of a *neu*-transformed cell line,[26–28] and immunization of mice with *neu* protein protects them from subsequent tumor challenge with the *neu*-transformed cell line.[14] Some authors have speculated that the use of antagonists for the unknown ligand could be useful in future chemotherapy.[85] Further review of this experimental evidence is beyond the scope of this article.

The c-*erb*B-2 activation most likely occurs at an early stage of neoplastic development. This hypothesis is supported by the presence of c-*erb*B-2 activation in both in situ and invasive breast carcinomas. In addition, studies of metastatic breast carcinomas usually demonstrate uniform c-*erb*B-2 activation at multiple sites in the same patient,[11,12,39,41,52] although c-*erb*B-2 activation has rarely been detected in metastatic lesions but not in the primary tumor.[57,66,107] Even more rarely, c-*erb*B-2 DNA amplification has been detected in a primary breast carcinoma but not in its lymph node metastasis.[5] In patients who have bilateral breast neoplasms, both lesions have similar patterns of c-*erb*B-2 activation, but only a few such cases have been studied.[11]

MECHANISMS OF c-*erb*B-2 ACTIVATION

The most common mechanism of c-*erb*B-2 activation is genomic DNA amplification, which almost always results in overproduction of c-*erb*B-2 mRNA and protein.[17,34,65,81] The c-*erb*B-2 amplification may stabilize the overproduction of mRNA or protein through unknown mechanisms. Human breast carcinomas with c-*erb*B-2 amplification contain 2 to 40 times more c-*erb*B-2 DNA[4,5] and 4 to 128 times more c-*erb*B-2 mRNA[34,90] than found in normal tissue. Most human breast carcinomas with c-*erb*B-2 amplification have 2 to 15 times more c-*erb*B-2 DNA. Tumors with greater amplification tend to have greater overproduction.[17,52,65] The non-mammary neoplasms that have been studied tend to have

similar levels of c-*erb*B-2 amplification or overproduction relative to the corresponding normal tissue.

The second most common mechanism of c-*erb*B-2 activation is overproduction of c-*erb*B-2 mRNA and protein without amplification of c-*erb*B-2 DNA.[81] The quantities of mRNA and protein usually are less than those in amplified cases and may approach the small quantities present in normal breast or other tissues.[17,50,52] The c-*erb*B-2 protein overproduction without mRNA overproduction or DNA amplification has been described in a few human breast carcinoma cell lines.[47]

Other rare mechanisms of c-*erb*B-2 activation have been reported. Translocations involving the c-*erb*B-2 gene have been described in a few mammary and gastric carcinomas, although some reported cases may represent restriction fragment length polymorphisms or incomplete restriction enzyme digestions that mimic translocations.[31,65,75,84,90,103] A single point mutation in the transmembrane portion of *neu* has been described in rat neuroblastomas induced by ethylnitrosurea.[9,55] The mutated *neu* protein has increased tyrosine kinase activity and aggregates at the cell membrane.[10,83,98] Although there has been speculation that some of the amplified c-*erb*B-2 genes may contain point mutations,[46] none has been detected in primary human neoplasms.[41,53,81]

TECHNIQUES FOR DETECTING c-*erb*B-2 ACTIVATION

Detection of c-*erb*B-2 DNA Amplification

Amplification of c-*erb*B-2 DNA is usually detected by DNA dot blot or Southern blot hybridization. In the dot blot method, the extracted DNA is placed directly on a nylon membrane and hybridized with a c-*erb*B-2 DNA probe. In the Southern blot method, the extracted DNA is treated with a restriction enzyme, and the fragments are separated by electrophoresis, transferred to a nylon membrane, and hybridized with a c-*erb*B-2 DNA probe. In both techniques, c-*erb*B-2 amplification is quantified by comparing the intensity (measured by densitometry) of the hybridization bands from the sample with those from control tissue.

Several technical problems may complicate the measurement of c-*erb*B-2 DNA amplification. First, the extracted tumor DNA may be excessively degraded or diluted by DNA from stromal cells.[81] Second, the c-*erb*B-2 DNA probe must be carefully chosen and labeled. For example, oligonucleotide c-*erb*B-2 probes may not be sensitive enough for measuring a low level of c-*erb*B-2 amplification, because diploid copy numbers can be difficult to detect (unpublished data). Third, the total amounts of DNA in the sample and control tissue must be compensated for, often with a probe to an unamplified gene. Many studies have used control probes to genes on chromosome 17, the location of c-*erb*B-2, to correct for possible alterations in chromosome number. Identical results, however, are obtained by using control probes to genes on other chromosomes,[5,65,80] with rare exception.[17] Studies using control probes to the beta-

globin gene must be interpreted with caution, because one allele of this gene is deleted occasionally in breast carcinomas.[3]

Amplification of c-*erb*B-2 DNA was assessed by using the polymerase chain reaction (PCR) in one recent study.[32] Oligoprimers for the c-*erb*B-2 gene and a control gene are added to the sample's DNA, and PCR is performed. If the sample contains more copies of c-*erb*B-2 DNA than of the control gene, the c-*erb*B-2 DNA is replicated preferentially.

Detection of c-*erb*B-2 mRNA Overproduction

Overproduction of c-*erb*B-2 mRNA usually is measured by RNA dot blot or Northern blot hybridization. Both techniques require extraction of RNA but otherwise are analogous to DNA dot blot and Southern blot hybridization. Use of PCR for detection of c-*erb*B-2 mRNA has been described in two recent abstracts.[89,102]

Overproduction of c-*erb*B-2 mRNA can be measured by in situ hybridization. Sections are mounted on glass slides, treated with protease, hybridized with a radiolabeled probe, washed, treated with nuclease to remove unbound probe, and developed for autoradiography. Silver grains are seen only over tumor cells that overproduce c-*erb*B-2 mRNA. Negative control probes are used.[65,86,106] Our experience indicates that these techniques are relatively insensitive for detecting c-*erb*B-2 mRNA overproduction in routinely processed tissue. Although the sensitivity may be increased by modifications that allow simultaneous detection of c-*erb*B-2 DNA and mRNA, in situ hybridization still is cumbersome and expensive (unpublished data).

All of the above c-*erb*B-2 mRNA detection techniques have several problems that make them more difficult to perform than techniques for detecting DNA amplification. One major problem is the rapid degradation of RNA in tissue that is not immediately frozen or fixed. In addition, during the detection procedure, RNA can be degraded by RNase, a ubiquitous enzyme, which must be eliminated meticulously from laboratory solutions. Third, control probes to genes that are uniformly expressed in the tissue of interest need to be carefully selected.

Detection of c-*erb*B-2 Protein Overproduction

The most accurate methods for detecting c-*erb*B-2 protein overproduction are the Western blot method and immunoprecipitation. Both techniques can document the binding specificity of various antibodies against c-*erb*B-2 protein. In Western blot studies, protein is extracted from the tissue, separated by electrophoresis (according to size), transferred to a membrane, and detected by using antibodies to c-*erb*B-2. In immunoprecipitation studies, antibodies against c-*erb*B-2 are added to a tumor lysate, and the resulting protein–antibody precipitate is separated by gel electrophoresis and stained for protein. Both Western blot and immunoprecipitation are useful research tools but currently are not practical for diagnostic pathology. Two recent abstracts have described an enzyme-linked immunosorbent assay (ELISA) for detection of c-*erb*B-2 protein.[19,45]

Overproduction of c-*erb*B-2 protein is most commonly assessed by various immunohistochemical techniques. These procedures often generate conflicting results, which are explained at least partially by three factors. First, various studies have used different polyclonal and monoclonal antibodies. Because some polyclonal antibodies recognize weak bands in addition to the c-*erb*B-2 protein band on Western blot or immunoprecipitation, the results of these studies should be interpreted with caution.[22,36,47,61] Even some monoclonal antibodies immunoprecipitate protein bands in addition to c-*erb*B-2 (p185).[30,59,86] Second, tissue fixation contributes to variability between studies. For example, some antibodies detect c-*erb*B-2 protein only in frozen tissue and do not react in fixed tissue. In general, formalin fixation diminishes the sensitivity of immunohistochemical methods and decreases the number of reactive cells.[81,86] When Bouin's fixative is used, there may be a higher percentage of positive cases.[22] Third, minimal criteria for interpreting immunohistochemical staining are generally lacking. Although there is general agreement that distinct crisp cytoplasmic membrane staining is diagnostic for c-*erb*B-2 activation in breast carcinoma, the number of positive cells and the staining intensity required to diagnose c-*erb*B-2 protein overproduction varies from study to study and from antibody to antibody. Degradation of c-*erb*B-2 protein is not a problem because it can be detected in intact form more than 24 hours after tumor resection without fixation or freezing.[84]

ACTIVATION OF c-*erb*B-2 IN BREAST LESIONS

Incidence of c-*erb*B-2 Activation

Most studies of c-*erb*B-2 oncogene activation do not specify histological subtypes of infiltrating breast carcinoma. Amplification of c-*erb*B-2 DNA was found in 19.1 percent (519 of 2715) of invasive carcinomas in 25 studies (Table 1), and c-*erb*B-2 mRNA or protein overproduction was detected in 20.9 percent (566 of 2714) of invasive carcinomas in 20 studies. Twelve studies have documented c-*erb*B-2 mRNA or protein overproduction in 15 percent (88 of 604) of carcinomas that lacked c-*erb*B-2 DNA amplification.

The incidence of c-*erb*B-2 activation in infiltrating breast carcinoma varies with the histological subtype. Approximately 22 percent (142 of 650) of infiltrating ductal carcinomas have c-*erb*B-2 activation, as expected from the above data. Other variants of breast carcinoma with frequent c-*erb*B-2 activation are inflammatory carcinoma (62 percent, 54 of 87), Paget's disease (82 percent, 9 of 11), and medullary carcinoma (22 percent, 5 of 23). In contrast, c-*erb*B-2 activation is infrequent in infiltrating lobular carcinoma (7 percent, 5 of 73) and tubular carcinoma (7 percent, 1 of 15).

The c-*erb*B-2 protein overproduction is present in 44 percent (44 of 100) of ductal carcinomas in situ and especially comedocarcinoma in situ (68 percent, 49 of 72). The micropapillary type of ductal carcinoma in situ also tends to have c-*erb*B-2 activation,[40,54,68] especially if larger cells are present. The greater fre-

TABLE 1. *c-erbB-2* ACTIVATION IN MALIGNANT HUMAN BREAST NEOPLASMS

Histological Diagnosis	c-*erbB*-2 DNA Amplification[a]	c-*erbB*-2 mRNA Overproduction	c-*erbB*-2 Protein Overproduction[b]
Carcinoma, not otherwise specified	146/526,[61] 52/310,[17] 52/291,[109] 28/176,[57] 17/157,[113] 22/141,[35] 14/136,[57] 12/122,[4] 19/103,[79] 15/95,[90] 15/86,[111] 17/73,[77] 16/66,[42] 6/61,[50] 11/57,[52] 10/57,[55] 13/51,[13] 8/49,[29] 10/38,[92] 12/36,[94] 1/25,[15] 7/24,[91] 7/15,[31] 7/10,[98] 2/10[107]	42/180,[59] 49/126,[35] 19/62,[65] 19/57,[50] 3/11,[90] 6/10,[96] 3/9[31]	118/728,[85b] 58/330,[17b] 47/313,[66] 17/195,[11] 32/191,[58] 31/185,[101] 34/102,[42] 24/53,[52b] 23/47,[13] 22/45,[6] 11/36,[94] 7/24,[91] 1/10[31]
Carcinoma, type not specified but lacking c-*erbB*-2 DNA amplification	—	16/136,[61] 14/73,[34] 8/16,[65] 0/8,[90] 1/4,[31] 0/3[96]	16/231,[17b] 18/135,[61] 13/35,[13] 14/29,[52b] 1/28,[92] 3/24,[94] 0/17[91]
Infiltrating ductal carcinoma	21/118,[92] 23/107,[34] 17/50,[44] 7/37[93] 14/53 (comedo-carcinoma)[16] 3/33 (tubuloductal carcinoma)[16]	35/95[34]	22/137,[40] 14/93,[69] 9/34[88]

Inflammatory carcinoma	33/60,[35] 3/6[52]	46/75[35]	5/6[52b]
Paget's disease	—	—	5/6,[40] 2/3,[54] 2/2[92]
Tubular carcinoma	0/5,[16] 0/1[93]	—	1/9[40]
Medullary carcinoma	2/4,[16] 0/1[34]	0/1[34]	1/12,[40] 1/3,[58] 1/2,[92] 0/1[39]
Mucinous carcinoma	0/1,[16] 0/1[93]	—	1/2[58]
Invasive papillary carcinoma	0/2[93]	—	—
Infiltrating lobular carcinoma	1/15,[16] 0/6[34]	1/5[34]	2/27,[92] 0/12,[40] 0/9,[39] 1/5[58]
Mammary fibrosarcoma	0/1[30]	—	—
"Benign cystosarcoma"	—	—	0/1[68]
Ductal CIS[c] with minimal invasion	3/5[92]	—	—
Ductal CIS	0/2[34]	1/2[34]	33/74,[40] 10/24[39]
Ductal CIS, solid or comedo type	—	—	20/33,[68] 19/29,[92] 10/10[54]
Ductal CIS, micropapillary type	—	—	10/10[68]
Ductal CIS, micropapillary or cribriform type	—	—	1(local)/14[54]
Ductal CIS, papillary or cribriform type	—	—	0/16,[92] 1/9,[68] 0/3[40]
Lobular CIS	—	—	0/16[40]

[a]Shown as number of cases with activation/number of cases studied; reference is given as a superscript.
[b]These protein studies used Western blots; the rest used immunohistochemical methods.
[c]CIS = carcinoma in situ.

quency of c-*erb*B-2 protein overproduction in comedocarcinoma in situ, compared with infiltrating ductal carcinoma, could be explained by the fact that many infiltrating ductal carcinomas arise from other types of intraductal carcinoma, which show c-*erb*B-2 activation infrequently. Others have speculated that carcinoma in situ with c-*erb*B-2 activation tends to regress or to lose c-*erb*B-2 activation during progression to invasion.[40,68,92] Infiltrating and in situ components of ductal carcinoma, however, usually are similar with respect to c-*erb*B-2 activation,[11,39] although some authors have noted more heterogeneity of the immunohistochemical staining pattern in invasive than in in situ carcinoma.[40,42,88] Activation of c-*erb*B-2 is infrequent in lobular carcinoma in situ. If lesions contain more than one histological pattern of carcinoma in situ, the c-*erb*B-2 protein overproduction tends to occur in the comedocarcinoma in situ but may include other areas of carcinoma in situ.[42,54,68] Overproduction of c-*erb*B-2 protein in ductal carcinoma in situ correlates with larger cell size and a periductal lymphoid infiltrate.[68]

Activation of c-*erb*B-2 has not been identified in benign breast lesions, including fibrocystic disease, fibroadenomas, and radial scars (Table 2). Strong membrane immunohistochemical reactivity for c-*erb*B-2 has not been described in atypical ductal hyperplasia, although weak accentuation of membrane staining has been noted infrequently.[39,42,54] In normal breast tissue, c-*erb*B-2 DNA is diploid, and c-*erb*B-2 is expressed at lower levels than in activated tumors.[34,35,65,88]

These preliminary data suggest that c-*erb*B-2 activation may not be useful for resolving many of the common problems in diagnostic surgical pathology. For example, c-*erb*B-2 activation is infrequent in tubular carcinoma and radial scars. In addition, because c-*erb*B-2 activation is unusual in atypical ductal hyperplasia, cribriform carcinoma in situ, and papillary carcinoma in situ, detection of c-*erb*B-2 activation in these lesions may not be helpful in their differential diagnosis. The histological features of comedocarcinoma in situ, which commonly overproduces c-*erb*B-2, are unlikely to be mistaken for those of benign lesions. Activation of

TABLE 2. c-*erb*B-2 ACTIVATION IN BENIGN HUMAN BREAST LESIONS

Histological Diagnosis	c-*erb*B-2 DNA Amplification[a]	c-*erb*B-2 mRNA Overproduction	c-*erb*B-2 Protein Overproduction
Fibrocystic disease	0/10[93]	—	0/32,[39] 0/9,[88] 0/8[88]
Atypical ductal hyperplasia	—	—	2(weak)/21,[54] 1(cytoplasmic)/13[39]
Benign ductal hyperplasia	—	—	0/12[39]
Sclerosing adenosis	—	—	0/4[39]
Fibroadenomas	0/16,[34] 0/6,[93] 0/2,[21] 0/1[91]	0/6,[35] 0/3[34]	0/21,[39] 0/10,[88] 0/8,[88] 0/3[42]
Radial scars	—	—	0/22[39]
Blunt duct adenosis	—	—	0/14[39]
"Breast mastosis"	—	0/3[35]	—

[a]Shown as number of cases with activation/number of cases studied; reference is given as a superscript.

c-*erb*B-2, however, does favor infiltrating ductal carcinoma over infiltrating lobular carcinoma. Further studies of these issues would be useful.

Correlation of c-*erb*B-2 Activation With Pathologic Prognostic Factors

Multiple studies have attempted to correlate c-*erb*B-2 activation with various pathologic prognostic factors (Table 3). Activation of c-*erb*B-2 was correlated with lymph node metastasis in 8 of 28 series, with higher histological grade in 6 of 17 series, and with higher stage in 4 of 14 series. Large tumor size was not associated with c-*erb*B-2 activation in most studies (11 of 14). Tetraploid DNA content and low proliferation, measured by Ki-67, have been suggested as prognostic factors and may correlate with c-*erb*B-2 activation.[6,7]

Correlation of c-*erb*B-2 Activation With Clinical Prognostic Factors

Various studies have attempted also to correlate c-*erb*B-2 activation with clinical features that may predict a poor outcome (Table 4). Activation of c-*erb*B-2 correlated with absence of estrogen receptors in 10 of 28 series and with absence of progesterone receptors in 6 of 18 series. In most studies, patient age did not correlate with c-*erb*B-2 activation, and, in the rest of the reports, c-*erb*B-2 activation was associated with either younger or older ages.

Correlation of c-*erb*B-2 Activation With Patient Outcome

Slamon et al[79,81] first showed that amplification of the c-*erb*B-2 oncogene independently predicts decreased survival of patients with breast carcinoma. The correlation of c-*erb*B-2 amplification with poor outcome was nearly as strong as the correlation of number of involved lymph nodes with poor outcome. Slamon et al also reported that c-*erb*B-2 amplification is an important prognostic indicator only in patients with lymph node metastasis.[79,81]

A large number of subsequent studies also attempted to correlate c-*erb*B-2 activation with prognosis (Table 5). In 12 series, there was a correlation between c-*erb*B-2 activation and tumor recurrence or decreased survival. In five of these series, the predictive value of c-*erb*B-2 activation was reported to be independent of other prognostic factors. In contrast, 18 series did not confirm the correlation of c-*erb*B-2 activation with recurrence or survival. Four possible explanations for this controversy are discussed below.

One problem is that c-*erb*B-2 amplification correlates with prognosis mainly in patients with lymph node metastasis. As summarized in Table 5, most studies of patients with axillary lymph node metastasis showed a correlation of c-*erb*B-2 activation with poor outcome. In contrast, most studies of patients without axillary metastasis have not demonstrated a correlation with patient outcome. Table 6 summarizes the studies in which all patients (with and without axillary metastasis) were considered as one group. There is a trend for studies with a higher percentage of metastatic cases to show an association between c-*erb*B-2 activation and poor outcome. Thus, most of the current evidence suggests that c-*erb*B-2 activation has prognostic value only in patients with metastasis to lymph nodes.

TABLE 3. CORRELATION OF c-*erb*B-2 ACTIVATION WITH PATHOLOGIC PROGNOSTIC FACTORS IN BREAST CARCINOMA

Prognostic Factor	*P*[a]	c-*erb*B-2 DNA Amplification[b]	c-*erb*B-2 mRNA Overproduction	c-*erb*B-2 Protein Overproduction[c]
Metastasis to axillary lymph nodes	<0.05	(118)[35] (105)[34] (49)[21]	(104)[35] (92)[34] (9)[31]	(350)[85c] (36)[13] (189)[92]
	0.05–0.15	(103)[79] (86)[79] (58)[111]	—	
	>0.15	(279)[17] (176)[87] (157)[113] (122)[4] (85)[90] (50)[62] (50)[44] (47)[15] (41)[93]	(50)[50]	(329)[17c] (261)[88] (195)[11] (185)[101] (102)[39] (50)[52b]
Larger size	<0.05	(280)[17]	—	(330)[17c] (189)[92]
	0.05–0.15	(86)[79]	—	—
	>0.15	(178)[87] (157)[113] (103)[79] (64)[77] (58)[111] (45)[21]	(51)[50]	(350)[85c] (185)[101] (34)[92]
Higher stage	<0.05	(300)[17] (64)[77] (58)[111]	—	(349)[17c]
	0.05–0.15	(56)[52]	—	—
	>0.15	(176)[87] (157)[113] (84)[90] (61)[50] (53)[21] (52)[57] (41)[93]	—	(102)[39] (56)[52b]
Higher histological grade	<0.05	(47)[13] (15)[31]	(53)[55]	(176)[101] (168)[11] (38)[13]
	0.05–0.15	—	—	—
	>0.15	(122)[4] (113)[34] (95)[90] (58)[111] (50)[44] (41)[93]	(86)[13] (65)[55]	(290)[86] (189)[92] (102)[39]

[a]A correlation is statistically significant at <0.05, equivocal at best between 0.05 and 0.15, and not statistically significant at >0.15.
[b]Numbers inside parentheses are the number of patients in an individual study; superscript is the reference. Some studies analyzed more than one group of patients.
[c]By Western blot method; all other protein studies used immunohistochemical methods.

TABLE 4. CORRELATION OF *c-erbB-2* ACTIVATION WITH CLINICAL PROGNOSTIC FACTORS IN BREAST CARCINOMA

Prognostic Factor	P[a]	c-*erbB*-2 DNA Amplification[b]	c-*erbB*-2 mRNA Overproduction	c-*erbB*-2 Protein Overproduction[c]
Absence of estrogen receptors	<0.05	(253)[109] (141)[35] (109)[34] (86)[79] (50)[44] (47)[13]	(104)[35]	(350)[85c] (330)[17a] (185)[101]
	0.05–0.15	—	—	—
	>0.15	(157)[113] (122)[4] (103)[79] (95)[90] (64)[77] (61)[50] (58)[111] (53)[21] (51)[62] (41)[93]	(180)[89] (62)[85] (62)[85] (57)[50]	(290)[88] (172)[11] (51)[62c] (38)[13]
Absence of progesterone receptors	<0.05	(253)[109] (141)[35] (109)[34] (50)[44]	—	(350)[85c] (306)[17a]
	0.05–0.15	(86)[79] (49)[62]	—	—
	>0.15	(157)[113] (122)[4] (103)[79] (64)[77]	(180)[89] (103)[95] (62)[85] (56)[35]	(90)[11] (49)[62a]
Age (menopausal status)	<0.05	—	—	(younger: 330)[17c] (older: 56)[52c]
	0.05–0.15	(younger: 86)[79]	—	—
	>0.15	(230)[17] (176)[67] (157)[113] (122)[4] (116)[34] (103)[79] (95)[90] (64)[77] (58)[111] (56)[52] (53)[21] (49)[13] (41)[93] (15)[21]	(62)[85]	(350)[85c] (290)[88] (189)[92] (162)[11] (45)[92]

[a]A correlation is statistically significant at <0.05, equivocal at best between 0.05 and 0.15, and not statistically significant at >0.15
[b]Numbers inside parentheses are the number of patients in an individual study; superscript is the reference. Some studies analyzed more than one group of patients.
[c]By Western blot method; all other protein studies used immunohistochemical methods.

TABLE 5. CORRELATION OF c-*erb*B-2 ACTIVATION WITH OUTCOME IN PATIENTS WITH BREAST CARCINOMA

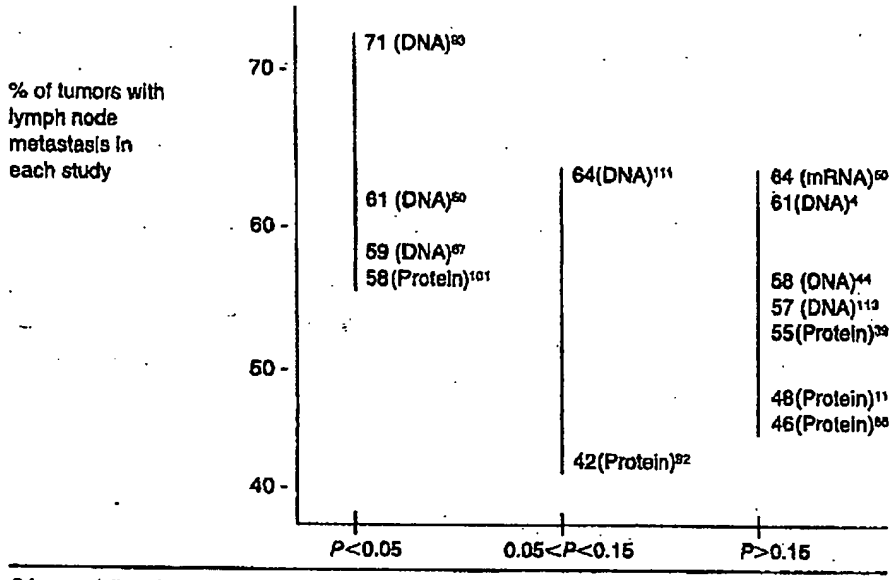
		Number of Patients				
P[a]	Type of c-*erb*B-2 Activation[b]	*Total*	*With Metastasis to Axillary Lymph Nodes*	*No Metastasis*	Statistical Analysis[c]	Reference
<0.05	DNA	176			M	87
<0.05	DNA	61			U	50
<0.05	DNA	57			U	65
<0.05	DNA	41			U	89
<0.05	mRNA	62			U	65
<0.05	Protein	102			M	101
<0.05	DNA		345		M	81
<0.05	DNA		120		U	17
<0.05	DNA		91		U	87
<0.05	DNA		86		M	79
<0.05	Protein-WB		350		M	85
<0.05	Protein		62	44	U	101
0.05–0.15	DNA	57			U	111
0.05–0.15	Protein	189			M	92
0.05–0.15	Protein		120		U	86
>0.15	DNA	130			U	113
>0.15	DNA	122			M	4
>0.15	DNA	50			U	44
>0.15	mRNA	57			U	50
>0.15	Protein	290			M	86
>0.15	Protein	195			U	11
>0.15	Protein	102			U	39
>0.15	Protein		137		U	17
>0.15	DNA			181	M	81
>0.15	DNA			159	U	17
>0.15	DNA			73	U	87
>0.15	Protein-WB			378	U	85
>0.15	Protein-WB			192	U	17
>0.15	Protein			141	U	86
>0.15	Protein			41	U	40

[a]The endpoints of these studies were tumor recurrence or decreased survival or both. Correlation between c-*erb*B-2 activation and a poorer patient outcome is statistically significant at <0.05, is of equivocal significance at 0.05 to 0.15, and is not significant at >0.15.

[b]Shown as variable measured. Letters "WB" indicate assay by Western blot; the other protein studies used immunohistochemical methods.

[c]M = multivariate statistical analysis; U = univariate statistical analysis.

TABLE 6. PERCENTAGE OF BREAST CARCINOMAS WITH METASTASIS COMPARED WITH PROGNOSTIC SIGNIFICANCE OF c-*erb*B-2 ACTIVATION



P for correlation of c-*erb*B-2 activation with patient outcome.
Each study's percentage of breast carcinomas with metastasis is compared with the correlation between c-*erb*B-2 activation and outcome. These data include only those studies that considered, as one group, all breast cancer patients, whether or not they had axillary metastasis. Superscripts are the references. In parentheses are the types of c-*erb*B-2 activation. *P* values are interpreted as in Table 3.

A second problem is that various types of breast carcinoma are grouped together in many survival studies. Because the current literature suggests that c-*erb*B-2 activation is infrequent in lobular carcinoma, studies that combine infiltrating ductal and lobular carcinomas may dilute the prognostic effect of c-*erb*B-2 activation in ductal tumors. In addition, most studies do not analyze inflammatory breast carcinoma separately. This condition frequently shows c-*erb*B-2 activation and has a worse prognosis than the usual mammary carcinoma, but it is an uncommon lesion.

A third potential problem is the paucity of studies that attempt to correlate c-*erb*B-2 activation with clinical outcome in subsets of breast carcinoma without metastasis. Two recent abstracts reported that in patients without lymph node metastasis who had various risk factors for recurrence (such as large tumor size and absence of estrogen receptors), c-*erb*B-2 overexpression predicted early recurrence.[23,67] In patients with ductal carcinoma in situ, one small study found no association between tumor recurrence and c-*erb*B-2 activation.[40]

A fourth problem is the lack of data regarding whether the prognosis correlates better with c-*erb*B-2 DNA amplification or with mRNA or protein overproduction. Most studies that find a correlation between c-*erb*B-2 activa-

tion and poor patient outcome measure c-*erb*B-2 DNA amplification (Table 5), and breast carcinoma patients with greater amplification of c-*erb*B-2 may have poorer survival.[79,81] Recent studies suggest that amplification has more prognostic power than overproduction,[17,34,35] but the clinical significance of c-*erb*B-2 overproduction without DNA amplification deserves further research.[17,52] Few studies have attempted to correlate patient outcome with c-*erb*B-2 mRNA overproduction, and many studies of c-*erb*B-2 protein overproduction use relatively less reliable methods such as immunohistochemical studies with polyclonal antibodies.

Comparison of c-*erb*B-2 Activation With Other Oncogenes in Breast Carcinoma

Other oncogenes that may have prognostic implications in human breast cancer are reviewed elsewhere.[71,106] This section will be restricted to a comparison between the clinical relevance of c-*erb*B-2 and these other oncogenes.

The c-*myc* gene is often activated in breast carcinomas, but c-*myc* activation generally has less prognostic importance than c-*erb*B-2 activation.[21,34,77,87,93] One study found a correlation between increased mRNAs of c-*erb*B-2 and c-*myc*, although other reports have not confirmed this.[34,106] Subsequent research, however, could demonstrate a subset of breast carcinomas in which c-*myc* has more prognostic importance than c-*erb*B-2.

The gene c-*erb*B-1 for the epidermal growth factor receptor (EGFR) is homologous with c-*erb*B-2 but is infrequently amplified in breast carcinomas.[79] Overproduction of EGFR, however, occurs more frequently than amplification and may correlate with a poor prognosis. In studies that have examined both c-*erb*B-2 and EGFR in the same tumor, c-*erb*B-2 has a stronger correlation with poor prognostic factors.[35,52] Studies have tended to show no correlation between amplification of c-*erb*B-2 and c-*erb*B-1 or overproduction of c-*erb*B-2 and EGFR, although at the molecular level EGFR mediates phosphorylation of c-*erb*B-2 protein.[51,52,61,68,100] Recent reviews describe EGFR in breast carcinoma.[43,100]

The genes c-*erb*A and *ear*-1 are homologous to the thyroid hormone receptor, and they are located adjacent to c-*erb*B-2 on chromosome 17. These genes are frequently coamplified with c-*erb*B-2 in breast carcinomas. The absence of c-*erb*A expression in breast carcinomas, however, is evidence against an important role for this gene in breast neoplasia.[90] Amplification of c-*erb*B-2 can occur without *ear*-1 amplification, and these tumors have a decreased survival that is similar to tumors with both c-*erb*B-2 and *ear*-1 amplification.[87] Consequently, c-*erb*B-2 amplification seems to be more important than amplification of c-*erb*A or *ear*-1.

Other genes also have been compared with c-*erb*B-2 activation in breast carcinomas. One study found a significant correlation between increased c-*erb*B-2 mRNA and increased mRNAs of *fos*, platelet-derived growth factor chain A, and Ki-*ras*.[106] Allelic deletion of c-Ha-*ras* may indicate a poorer prognosis in breast carcinoma,[21] but it has not been compared with c-*erb*B-2 activation. Some studies have suggested a correlation between advanced stage or recurrence of breast carcinoma and activation of any one of several oncogenes.[21,113]

ACTIVATION OF c-*erb*B-2 IN NON-MAMMARY TISSUES

Incidence of c-*erb*B-2 Activation in Non-Mammary Tissues

Table 7 summarizes the normal tissues in which c-*erb*B-2 expression has been detected, usually with immunohistochemical methods using polyclonal anti-

TABLE 7. PRESENCE OR ABSENCE OF c-*erb*B-2 mRNA OR c-*erb*B-2 PROTEIN IN NORMAL HUMAN TISSUES

Tissues With c-*erb*B-2 mRNA	Tissues Producing c-*erb*B-2 Protein[a]	Tissues Lacking c-*erb*B-2 mRNA	Tissues Lacking c-*erb*B-2 Protein
Skin[24]	Epidermis[56]		
	External root sheath[55]		
	Eccrine sweat gland[56]		
	Fetal oral mucosa[52]		Postnatal oral mucosa[52]
	Fetal esophagus[52]		Postnatal esophagus[52]
Stomach[24]	Stomach[22,52]		
	Fetal intestine[52a]		
Jejunum[24]	Small intestine[22,52]		
Colon[24]	Colon[22,52]		
Kidney[24]	Fetal kidney[52a]	Kidneys[104]	Glomerulus[52]
			Postnatal Bowman's capsule[52]
	Fetal proximal tubule[52]		Postnatal proximal tubule[52]
	Distal tubule[52]		
	Fetal collecting duct[52]		Postnatal collecting duct[52]
	Fetal renal pelvis[52]		Postnatal renal pelvis[52]
	Fetal ureter[52]		Postnatal fetal ureter[52]
Liver[24]	Hepatocytes[22]		Liver[52,95]
	Pancreatic acini[22]		
	Pancreatic ducts[22,52]		
	Endocrine cells of islets of Langerhans[22]		Pancreatic islets[52]
Lung[24]	Fetal trachea[52]		Postnatal trachea[52]
	Fetal bronchioles[52]		Postnatal bronchioles[52]
	Bronchioles[59]		
			Postnatal alveoli[52,59]
Fetal brain[24]			Postnatal brain[52]
	Fetal ganglion cells[52]		Postnatal ganglion cells[52]
Thyroid[1]			
Uterus[24]			
	Ovary[12]		
	Blood vessels[42]		Endothelium[52]
Placenta[24]			
			Adrenocortical cells[52]
			Postnatal thymus[52]
			Fibroblasts[52]
			Smooth muscle cells[52]
			Cardiac muscle cells[52]

[a]This protein study used Western blots; the rest used immunohistochemical methods.

bodies. Only a few studies have been performed, and some of these do not demonstrate convincing cell membrane reactivity in the published photographs. The interpretations in these studies, however, are listed, with the caveat that these findings should be confirmed by immunoprecipitation or Western or RNA blots. Production of c-*erb*B-2 has been identified in normal epithelium of the gastrointestinal tract and skin. Discrepancies regarding c-*erb*B-2 protein in other tissues could be due, at least in part, to differences in techniques.

The data on c-*erb*B-2 activation in various non-mammary neoplasms should be interpreted with caution, because only small numbers of tumors have been studied, usually by immunohistochemical methods using polyclonal antibodies. Studies using cell lines have been excluded, because cell culture can induce amplification and overexpression of other genes, although this has not been documented for c-*erb*B-2.

Activation of c-*erb*B-2 has been identified in 32 percent (64 of 203) of ovarian carcinomas in eight studies (Table 8). One abstract[45] stated that ovarian carcinomas contained significantly more c-*erb*B-2 protein than ovarian non-epithelial malignancies. Another report[81] showed that 12 percent of ovarian carcinomas had c-*erb*B-2 overproduction without amplification.

Activation of c-*erb*B-2 has been identified in 20 percent (40 of 198) of gastric adenocarcinomas in seven studies, including 33 percent (21 of 64) of

TABLE 8. c-*erb*B-2 ACTIVATION IN HUMAN GYNECOLOGIC TUMORS[a]

Tumor Type	c-*erb*B-2 DNA Amplification	c-*erb*B-2 mRNA Over-production	c-*erb*B-2 Protein Over-production
Ovary—carcinoma, not otherwise specified	31/120,[81] 1/11,[57] 0/5,[107] 0/5,[84] 0/3,[112] 0/2,[72] 0/1[110]	23/67[81]	23/73,[12] 36/72[81]
Ovary—serous (papillary) carcinoma	2/7,[110] 1/7,[112] 0/6[72]	—	—
Ovary—endometrioid carcinoma	0/3[110]	—	—
Ovary—mucinous carcinoma	1/2,[110] 0/1[72]	—	—
Ovary—clear cell carcinoma	0/2,[112] 0/1[72]	—	—
Ovary—mixed epithelial carcinoma	0/2[72]	—	—
Ovary—endometrioid borderline tumor	0/1[72]	—	—
Ovary—mucinous borderline tumor	0/3[72]	—	—
Ovary—serous cystadenoma	0/4[72]	—	—
Ovary—mucinous cystadenoma	0/2[72]	—	—
Ovary—sclerosing stromal tumor	0/1[72]	—	—
Ovary—fibrothecoma	0/1[72]	—	—
Uterus—endometrial adenocarcinoma	0/4,[84] 0/1[110]	—	—

[a]Shown as number of cases with amplification (or overproduction)/total number of cases studied; reference is given as superscript. All protein studies used immunohistochemical methods.

intestinal or tubular subtypes and 9 percent (4 of 47) of diffuse or signet ring cell subtypes (Table 9). Activation of c-*erb*B-2 has been detected in 2 percent (6 of 281) of colorectal carcinomas, although an additional immunohistochemical study detected c-*erb*B-2 protein in seven of eight tissues fixed in Bouin's solution. One study found greater immunohistochemical reactivity for c-*erb*B-2 protein in colonic adenomatous polyps than in the adjacent normal epithelium, using Bouin's fixative. Lesions with anaplastic features and progression to invasive carcinoma tended to show decreased immunohistochemical reactivity for c-*erb*B-2 protein.[22] Hepatocellular carcinomas (12 of 14 cases) and cholangiocarcinomas (46 of 63 cases) reacted with antibodies against c-*erb*B-2 in one study, but some of these "positive" cases showed only diffuse cytoplasmic staining, which

TABLE 9. c-*erb*B-2 ACTIVATION IN HUMAN GASTROINTESTINAL TUMORS[a]

Tumor Type	c-*erb*B-2 DNA Amplification	c-*erb*B-2 Protein Overproduction
Esophagus—squamous cell carcinoma	0/1[107]	0/1[81]
Stomach—carcinoma, poorly differentiated	0/22[108]	—
Stomach—adenocarcinoma	2/24,[84] 2/9,[107] 2/8,[111] 2/8,[57] 0/1[108]	4/27,[29] 3/10[51]
Stomach—carcinoma, intestinal or tubular type	5/10[108]	16/54[29]
Stomach—carcinoma, diffuse or signet ring cell type	0/2[108]	4/45[29]
Colorectum—carcinoma	2/49,[84] 1/45,[111] 1/45,[57] 1/45,[60] 0/40,[61] 0/32,[107] 0/3[82]	1/22,[58] 7/8[22b]
Colon—villous adenoma	0/1[60]	—
Colon—tubulovillous adenoma	0/5[60]	—
Colon—tubular adenoma	0/7[60]	19/19[22b]
Colon—hyperplastic polyp	0/1[60]	—
Intestine—leiomyosarcoma	—	0/1[61]
Hepatocellular carcinoma	0/12[111]	12/14,[85] 0/2[51]
Hepatoblastoma	0/1[57]	—
Cholangiocarcinoma	—	46/63[85]
Pancreas—adenocarcinoma	—	2/80,[41c] 0/2[51]
Pancreas—acinar carcinoma	—	0/1[41]
Pancreas—clear cell carcinoma	—	0/2[41]
Pancreas—large cell carcinoma	—	0/3[41]
Pancreas—signet ring carcinoma	—	0/1[41]
Pancreas—chronic inflammation	—	0/14[41c]

[a]Shown as number of cases with amplification (or overproduction)/total number of cases studied; reference is given as superscript. All protein studies used immunohistochemical methods. No studies analyzed for c-*erb*B-2 mRNA.
[b]Tissues fixed in Bouin's solution.
[c]Only cases with distinct membrane staining are interpreted as showing c-*erb*B-2 overproduction.

TABLE 10. c-*erb*B-2 ACTIVATION IN HUMAN PULMONARY TUMORS[a]

Tumor Type	c-*erb*B-2 DNA Amplification	c-*erb*B-2 Protein Overproduction
Non-small cell carcinoma	2/60,[75] 0/60[81]	1/84[59]
Epidermoid carcinoma	0/13,[82] 0/10,[57] 0/6[20]	3/5[99]
Adenocarcinoma	0/21,[82] 1/13,[20] 0/7,[111] 0/7,[57] 0/3[107]	4/12[99]
Large cell carcinoma	0/9,[82] 0/6[20]	—
Small cell carcinoma	—	0/26,[58] 0/3[99]
Carcinoid tumor	0/1[82]	0/3[99]

[a]Shown as number of cases with amplification (or overproduction)/total number of cases studied; reference is given as superscript. All protein studies used immunohistochemical methods. No studies analyzed for c-*erb*B-2 mRNA.

does not indicate c-*erb*B-2 activation in breast neoplasms.[95] Also, some pancreatic carcinomas and chronic pancreatitis tissue had cytoplasmic immunohistochemical reactivity for c-*erb*B-2 protein, in addition to the rare case of pancreatic adenocarcinoma with distinct cell membrane staining.[41]

Tables 10 through 14 summarize the studies of c-*erb*B-2 activation in other neoplasms. The c-*erb*B-2 oncogene is not activated in most of these tumors. Activation of c-*erb*B-2 has been detected in 1 percent (4 of 299) of pulmonary non-small cell carcinomas in nine studies, although one additional report[99] found c-*erb*B-2 protein overproduction in 41 percent (7 of 17). Renal cell carcinoma had c-*erb*B-2 activation in 7 percent (2 of 30) in four studies. Overproduction of c-*erb*B-2 protein was described in one transitional cell carcinoma of the urinary bladder, a grade 2 papillary lesion.[58] Squamous cell carcinoma and basal cell carcinoma of the skin may contain c-*erb*B-2 protein, but it is not clear

TABLE 11. c-*erb*B-2 ACTIVATION IN HUMAN HEMATOLOGIC PROLIFERATIONS[a]

Tumor Type	c-*erb*B-2 DNA Amplification	c-*erb*B-2 mRNA Overproduction	c-*erb*B-2 Protein Overproduction
Hematologic malignancies	0/23[111]	—	—
Malignant lymphoma	0/9,[57] 0/3[107]	0/1[1]	0/15[61]
Acute leukemia	0/14[57]	—	—
Acute lymphoblastic leukemia	0/1[107]	—	—
Acute myeloblastic leukemia	0/3[107]	—	—
Chronic leukemia	0/16[57]	—	—
Chronic lymphocytic leukemia	0/6[107]	—	—
Chronic myelogenous leukemia	0/8[107]	—	—
Myeloproliferative disorder	0/1[57]	—	—

[a]Shown as number of cases with amplification (or overproduction)/total number of cases studied; reference is given as superscript. All protein studies used immunohistochemical methods.

TABLE 12. c-*erb*B-2 ACTIVATION IN HUMAN TUMORS OF SOFT TISSUE AND BONE[a]

Tumor Type	c-*erb*B-2 DNA Amplification
Sarcoma	0/10,[111] 0/8[57]
Malignant fibrous histiocytoma	0/1[107]
Liposarcoma	0/3[107]
Pleomorphic sarcoma	0/1[107]
Rhabdomyosarcoma	0/1[107]
Osteogenic sarcoma	0/2,[107] 0/2[57]
Chondrosarcoma	0/1[107]
Ewing's sarcoma	0/1[57]
Schwannoma	0/1[57]

[a]Shown as number of cases with amplification (or overproduction)/total number of cases studied; reference is given as superscript. No studies analyzed for c-*erb*B-2 mRNA or c-*erb*B-2 protein.

whether the protein level is increased over that of normal skin.[56] Thyroid carcinomas and adenomas can have low levels of increased c-*erb*B-2 mRNA. One abstract described low-level c-*erb*B-2 DNA amplification in one of ten salivary gland pleomorphic adenomas.[48]

Correlation of c-*erb*B-2 Activation With Patient Outcome

Very few studies have attempted to correlate c-*erb*B-2 activation in non-mammary tumors with outcome. Slamon et al[81] showed that c-*erb*B-2 amplification or overexpression in ovarian carcinomas correlates with decreased survival, especially when marked activation is present. However, they did not report the stage, histological grade, or histological subtype of these neoplasms. Another study of stages III and IV ovarian carcinomas found a correlation between decreased survival and c-*erb*B-2 protein overproduction, but not between survival and histological grade.[18] One abstract stated that c-*erb*B-2 protein overproduction in 10 of 16 pulmonary adenocarcinomas correlated with decreased disease-free interval.[70] Another abstract described a tendency for immunohisto-

TABLE 13. c-*erb*B-2 ACTIVATION IN HUMAN TUMORS OF THE URINARY TRACT[a]

Tumor Type	c-*erb*B-2 DNA Amplification	c-*erb*B-2 mRNA Overproduction	c-*erb*B-2 Protein Overproduction
Kidney—renal cell carcinoma	1/5,[57] 1/4,[107] 0/5[84]	0/16[104]	—
Wilms' tumor	0/4[57]	—	—
Prostate—adenocarcinoma	—	—	0/23[58]
Urinary bladder—carcinoma	—	—	1/48[58]

[a]Shown as number of cases with amplification (or overproduction)/total number of cases studied; reference is given as superscript. All protein studies used immunohistochemical methods.

TABLE 14. c-*erb*B-2 ACTIVATION IN MISCELLANEOUS HUMAN TUMORS[a]

Tumor Type	c-*erb*B-2 DNA Amplification	c-*erb*B-2 mRNA Overproduction	c-*erb*B-2 Protein Overproduction
Skin—malignant melanoma	—	—	0/10[58]
Skin, head and neck—squamous cell carcinoma	0/7[107]	—	—
Site not stated—squamous cell carcinoma	0/8,[57] 0/2[75]	—	—
Salivary gland—adenocarcinoma	1/1[75]	—	—
Parotid gland—adenoid cystic carcinoma	—	—	0/1[51]
Thyroid—anaplastic carcinoma	0/1[1]	0/1[1]	—
Thyroid—papillary carcinoma	0/5[1]	3(low levels)/5[1]	—
Thyroid—adenocarcinoma	0/1[84]	—	—
Thyroid—adenoma	0/2[1]	1(low levels)/2[1]	—
Neuroblastoma	0/35,[81] 0/9,[57] 0/1[75]	—	—
Meningioma	0/2[57]	—	—

[a]Shown as number of cases with amplification (or overproduction)/total number of cases studied; reference is given as superscript. All protein studies used immunohistochemical methods.

chemical reactivity for c-*erb*B-2 protein to correlate with higher grades of prostatic adenocarcinoma.[97] Additional prognostic studies of ovarian carcinomas and other neoplasms are needed.

SUMMARY

Activation of the c-*erb*B-2 oncogene can occur by amplification of c-*erb*B-2 DNA and by overproduction of c-*erb*B-2 mRNA and c-*erb*B-2 protein. Approximately 20 percent of breast carcinomas show evidence of c-*erb*B-2 activation, which correlates with a poor prognosis primarily in patients with metastasis to axillary lymph nodes. Studies that have attempted to correlate c-*erb*B-2 activation with other prognostic factors in breast carcinoma have reported conflicting conclusions. The pathologic and clinical significance of c-*erb*B-2 activation in other neoplasms is unclear and should be assessed by additional studies.

REFERENCES

1. Aasland R, Lillehaug JR, Male R, et al. Expression of oncogenes in thyroid tumors: Coexpression of c-*erb*B2/*neu* and c-*erb*B. *Br J Cancer.* 57:358, 1988
2. Akiyama T, Sudo C, Ogawara H, et al. The product of the human c-*erb*B-2 gene: A 185-kilodalton glycoprotein with tyrosine kinase activity. *Science.* 232:1644, 1986

3. Ali IU, Lidereau R, Theillet C, Callahan R. Reduction to homozygosity of genes on chromosome 11 in human breast neoplasia. *Science.* 238:185, 1987
4. Ali IU, Campbell G, Lidereau R, Callahan R. Amplification of c-*erb*B-2 and aggressive human breast tumors. *Science.* 240:1795, 1988
5. Ali IU, Campbell G, Lidereau R, Callahan R. Lack of evidence for the prognostic significance of c-*erb*B-2 amplification in human breast carcinoma. *Oncogene Res.* 3:139, 1988
6. Bacus SS, Bacus JW, Slamon DJ, Press MF. HER-2/*neu* oncogene expression and DNA ploidy analysis in breast cancer. *Arch Pathol Lab Med.* 114:164, 1990
7. Bacus SS, Ruby SG, Weinberg DS, et al. HER-2/*neu* oncogene expression and proliferation in breast cancers. *Am J Pathol.* 137:103, 1990
8. Bargmann CI, Hung MC, Weinberg RA. The *neu* oncogene encodes an epidermal growth factor receptor-related protein. *Nature.* 319:226, 1986
9. Bargmann CI, Hung MC, Weinberg RA. Multiple independent activations of the *neu* oncogene by a point mutation altering the transmembrane domain of p185. *Cell.* 45:649, 1986
10. Bargmann CI, Weinberg RA. Oncogenic activation of the *neu*-encoded receptor protein by point mutation and deletion. *EMBO J.* 7:2043, 1988
11. Barnes DM, Lammie GA, Millis RR, et al. An immunohistochemical evaluation of c-*erb*B-2 expression in human breast carcinoma. *Br J Cancer.* 58:448, 1988
12. Berchuck A, Kamel A, Whitaker R, et al. Overexpression of *HER-2/neu* is associated with poor survival in advanced epithelial ovarian cancer. *Cancer Res.* 50:4087, 1990
13. Berger MS, Locher GW, Saurer S, et al. Correlation of c-*erb*B-2 gene amplification and protein expression in human breast carcinoma with nodal status and nuclear grading. *Cancer Res.* 48:1238, 1988
14. Bernards R, Destree A, McKenzie S, et al. Effective tumor immunotherapy directed against an oncogene-encoded product using a vaccinia virus vector. *Proc Natl Acad Sci USA.* 84:6854, 1987
15. Biunno I, Pozzi MR, Pierotti MA, et al. Structure and expression of oncogenes in surgical specimens of human breast carcinomas. *Br J Cancer.* 57:464, 1988
16. Borg Å, Linell F, Idvall I, et al. *Her2/neu* amplification and comedo type breast carcinoma. *Lancet.* 1:1268, 1989
17. Borg Å, Tandon AK, Sigurdsson H, et al. *HER-2/neu* amplification predicts poor survival in node-positive breast cancer. *Cancer Res.* 50:4332, 1990
18. Bouchard L, Lamarre L, Tremblay PJ, Jolicoeur P. Stochastic appearance of mammary tumors in transgenic mice carrying the MMTV/c-*neu* oncogene. *Cell.* 57:931, 1989
19. Carney WP, Retos C, Petit D, et al. Quantitation of the *neu* oncogene protein using a monoclonal antibody based immunoassay (abstract). *Mod Pathol.* 3:15A, 1990
20. Cline MJ, Battifora H. Abnormalities of protooncogenes in non-small cell lung cancer: Correlations with tumor type and clinical characteristics. *Cancer.* 60:2669, 1987
21. Cline MJ, Battifora H, Yokota J. Proto-oncogene abnormalities in human breast cancer: Correlations with anatomic features and clinical course of disease. *J Clin Oncol.* 5:999, 1987
22. Cohen JA, Weiner DB, More KF, et al. Expression pattern of the *neu* (NGL) gene-encoded growth factor receptor protein ($p185^{neu}$) in normal and transformed epithelial tissues of the digestive tract. *Oncogene.* 4:81, 1989
23. Colnaghi MI, Miotti S, Andreola S, et al. New prognostic factors in breast cancer (abstract). *Am Assoc Cancer Res Ann Meeting.* 30:230A, 1989

24. Coussens L, Yang-Feng TL, Liao YC, et al. Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with *neu* oncogene. *Science.* 230:1132, 1985
25. Di Fiore PP, Pierce JH, Kraus MH, et al. *erbB*-2 is a potent oncogene when overexpressed in NIH/3T3 cells. *Science.* 237:178, 1987
26. Drebin JA, Link VC, Weinberg RA, Greene MI. Inhibition of tumor growth by a monoclonal antibody reactive with an oncogene-encoded tumor antigen. *Proc Natl Acad Sci USA.* 83:9129, 1986
27. Drebin JA, Link VC, Greene MI. Monoclonal antibodies reactive with distinct domains of the *neu* oncogene-encoded p185 molecule exert synergistic anti-tumor effects in vivo. *Oncogene.* 2:273, 1988
28. Drebin JA, Link VC, Greene MI. Monoclonal antibodies specific for the *neu* oncogene product directly mediate anti-tumor effects in vivo. *Oncogene.* 2:387, 1988
29. Falck VG, Gullick WJ. c-*erbB*-2 oncogene product staining in gastric adenocarcinoma. An immunohistochemical study. *J Pathol.* 159:107, 1989
30. Fendly BM, Winget M, Hudziak RM, et al. Characterization of murine monoclonal antibodies reactive to either the human epidermal growth factor receptor or *HER2/neu* gene product. *Cancer Res.* 50:1550, 1990
31. Fontaine J, Tesseraux M, Klein V, et al. Gene amplification and expression of the *neu* (c-*erbB*-2) sequence in human mammary carcinoma. *Oncology.* 45:360, 1988
32. Frye RA, Benz CC, Liu E. Detection of amplified oncogenes by differential polymerase chain reaction. *Oncogene.* 4:1153, 1989
33. Fukushige SI, Matsubara KI, Yoshida M, et al. Localization of a novel v-*erbB*-related gene, c-*erbB*-2, on human chromosome 17 and its amplification in a gastric cancer cell line. *Mol Cell Biol.* 6:955, 1986
34. Guerin M, Barrois M, Terrier MJ, et al. Overexpression of either c-*myc* or c-*erbB*-2/*neu* proto-oncogenes in human breast carcinomas: Correlation with poor prognosis. *Oncogene Res.* 3:21, 1988
35. Guerin M, Gabillot M, Mathieu MC, et al. Structure and expression of c-*erbB*-2 and EGF receptor genes in inflammatory and non-inflammatory breast cancer: Prognostic significance. *Int J Cancer.* 43:201, 1989
36. Gullick WJ, Berger MS, Bennett PLP, et al. Expression of the c-*erbB*-2 protein in normal and transformed cells. *Int J Cancer.* 40:246, 1987
37. Gullick WJ, Venter DJ. The c-*erbB*2 gene and its expression in human cancers. In: Waxman J, Sikora K, eds. *The Molecular Biology of Cancer.* Boston, Blackwell Sci Publ; 1989: 38–53
38. Gullick WJ. Expression of the c-*erbB*-2 proto-oncogene protein in human breast cancer. *Recent Results Cancer Res.* 113:51, 1989
39. Gusterson BA, Machin LG, Gullick WJ, et al. c-*erbB*-2 expression in benign and malignant breast disease. *Br J Cancer.* 58:453, 1988
40. Gusterson BA, Machin LG, Gullick WJ, et al. Immunohistochemical distribution of c-*erbB*-2 in infiltrating and in situ breast cancer. *Int J Cancer.* 42:842, 1988
41. Hall PA, Hughes CM, Staddon SL, et al. The c-*erbB*-2 proto-oncogene in human pancreatic cancer. *J Pathol.* 161:195, 1990
42. Hanna W, Kahn HJ, Andrulis I, Pawson T. Distribution and patterns of staining of *neu* oncogene product in benign and malignant breast diseases. *Mod Pathol.* 3:455, 1990
43. Harris AL, Nicholson S. Epidermal growth factor receptors in human breast cancer.

In: Lippman ME, Dickson RB, eds. *Breast Cancer: Cellular and Molecular Biology.* Boston, Kluwer Academic Publ; 1988: 93–118

44. Heintz NH, Leslie KO, Rogers LA, Howard PL. Amplification of the c-*erb*B-2 oncogene and prognosis of breast adenocarcinoma. *Arch Pathol Lab Med.* 114:160, 1990
45. Huettner P, Carney W, Delellis R, et al. Quantification of the *neu* oncogene product in ovarian neoplasms (abstract). *Mod Pathol.* 3:46A, 1990
46. Hung MC, Yan DH, Zhao X. Amplification of the proto-*neu* oncogene facilitates oncogenic activation by a single point mutation. *Proc Natl Acad Sci USA.* 86:2545, 1989
47. Hynes NE, Gerber HA, Saurer S, Groner B. Overexpression of the c-*erb*B-2 protein in human breast tumor cell lines. *J Cell Biochem.* 39:167, 1989
48. Kahn HJ, Hanna W, Auger M, Andreulis I. Expression and amplification of *neu* oncogene in pleomorphic adenoma of salivary glands (abstract). *Mod Pathol.* 3:50A, 1990
49. King CR, Kraus MH, Aaronson SA. Amplification of a novel v-*erb*B-related gene in a human mammary carcinoma. *Science.* 229:974, 1985
50. King CR, Swain SM, Porter L, et al. Heterogeneous expression of *erb*B-2 messenger RNA in human breast cancer. *Cancer Res.* 49:4185, 1989
51. Kokai Y, Dobashi K, Weiner DB, et al. Phosphorylation process induced by epidermal growth factor receptor alters the oncogenic and cellular *neu* (*NGL*) gene products. *Proc Natl Acad Sci USA.* 85:5389, 1988
52. Lacroix H, Iglehart JD, Skinner MA, Kraus MH. Overexpression of *erb*B-2 or EGF receptor proteins present in early stage mammary carcinoma is detected simultaneously in matched primary tumors and regional metastases. *Oncogene.* 4:145, 1989
53. Lemoine NR, Staddon S, Dickson C, et al. Absence of activating transmembrane mutations in the c-*erb*B-2 proto-oncogene in human breast cancer. *Oncogene.* 5:237, 1990
54. Lodato RF, Maguire HC, Greene MJ, et al. Immunohistochemical evaluation of c-*erb*B-2 oncogene expression in ductal carcinoma in situ and atypical ductal hyperplasia of the breast. *Mod Pathol.* 3:449, 1990
55. Maguire HC, Greene MI. The *neu* (c-*erb*B-2) oncogene. *Semin Oncol.* 16:148, 1989
56. Maguire HC, Jaworsky C, Cohen JA, et al. Distribution of *neu* (c-*erb*B-2) protein in human skin. *J Invest Dermatol.* 89:786, 1989
57. Masuda H, Battifora H, Yokota J, et al. Specificity of proto-oncogene amplification in human malignant diseases. *Mol Biol Med.* 4:213, 1987
58. McCann A, Dervan PA, Johnston PA, et al. c-*erb*B-2 oncoprotein expression in primary human tumors. *Cancer.* 65:88, 1990
59. McKenzie SJ, Marks PJ, Lam T, et al. Generation and characterization of monoclonal antibodies specific for the human *neu* oncogene product, p185. *Oncogene.* 4:543, 1989
60. Meltzer SJ, Ahnen DJ, Battifora H, et al. Protooncogene abnormalities in colon cancers and adenomatous polyps. *Gastroenterology.* 92:1174, 1987
61. Mori S, Akiyama T, Morishita Y, et al. Light and electron microscopical demonstration of c-*erb*B-2 gene product-like immunoreactivity in human malignant tumors. *Virchows Arch [B].* 54:8, 1987
62. Mori S, Akiyama T, Yamada Y, et al. C-*erb*B-2 gene product, a membrane protein commonly expressed in human fetal epithelial cells. *Lab Invest.* 61:93, 1989
63. Muller WJ, Sinn E, Pattengale PK, et al. Single-step induction of mammary

adenocarcinoma in transgenic mice bearing the activated c-*neu* oncogene. *Cell.* 54:105, 1988

64. Ong G, Gullick W, Sikora K. Oncoprotein stability after tumor resection. *Br J Cancer.* 61:538, 1990
65. Parks HC, Lillycrop K, Howell A, Craig RK. C-*erb*B2 mRNA expression in human breast tumors: Comparison with c-*erb*B2 DNA amplification and correlation with prognosis. *Br J Cancer.* 61:39, 1990
66. Popescu NC, King CR, Kraus MH. Localization of the *erb*B-2 gene on normal and rearranged chromosomes 17 to bands q12–21.32. *Genomics.* 4:362, 1989
67. Press MF, Pike MC, Paterson MC, et al. Overexpression of *HER-2/neu*, proto-oncogene in node negative breast cancer: Correlation with increased risk of early recurrent disease (abstract). *Mod Pathol.* 3:80A, 1990
68. Ramachandra S, Machin L, Ashley S, et al. Immunohistochemical distribution of c-*erb*B-2 in in situ breast carcinoma: A detailed morphological analysis. *J Pathol.* 161:7, 1990
69. Rio MC, Bellocq JP, Gairard B, et al. Specific expression of the pS2 gene in subclasses of breast cancers in comparison with expression of the estrogen and progesterone receptors and the oncogene *ERBB2. Proc Natl Acad Sci USA.* 84:9243, 1987
70. Robinson R, Kern J, Weiner D, et al. $p185^{neu}$ expression in human lung non-small cell carcinomas: An immunohistochemical study with clinicopathologic correlation (abstract). *Mod Pathol.* 3:85A, 1990
71. Rochlitz CF, Benz CC. Oncogenes in human solid tumors. In: Benz C, Liu E, eds. *Oncogenes.* Boston, Kluwer Academic Publ; 1989: 199–240
72. Sasano H, Garret CT, Wilkinson DS, et al. Protooncogene amplification and tumor ploidy in human ovarian neoplasms. *Hum Pathol.* 21:382, 1990
73. Schechter AL, Stern DF, Vaidyanathan L, et al. The *neu* oncogene: An *erb*-B-related gene encoding a 185,000-M_r tumor antigen. *Nature.* 312:513, 1984
74. Schechter AL, Hung MC, Vaidyanathan L, et al. The *neu* gene: An *erb*B-homologous gene distinct from and unlinked to the gene encoding the EGF receptor. *Science.* 229:976, 1985
75. Schneider PM, Hung MC, Chiocca SM, et al. Differential expression of the c-*erb*B-2 gene in human small cell and non-small cell lung cancer. *Cancer Res.* 49:4968, 1989
76. Semba K, Kamata N, Toyoshima K, Yamamoto T. A v-*erb*B-related protooncogene, c-*erb*B-2, is distinct from the c-*erb*B-1/epidermal growth factor-receptor gene and is amplified in a human salivary gland adenocarcinoma. *Proc Natl Acad Sci USA.* 82:6497, 1985
77. Seshadri R, Matthews C, Dobrovic A, Horsfall DJ. The significance of oncogene amplification in primary breast cancer. *Int J Cancer.* 43:270, 1989
78. Shih C, Padhy LC, Murray M, Weinberg RA. Transforming genes of carcinomas and neuroblastomas introduced into mouse fibroblasts. *Nature.* 290:261, 1981
79. Slamon DJ, Clark GM, Wong SG, et al. Human breast cancer: Correlation of relapse and survival with amplification of the *HER-2/neu* oncogene. *Science.* 235:177, 1987
80. Slamon DJ, Clark GM. Amplification of c-*erb*B-2 and aggressive human breast tumors. *Science.* 240:1795, 1988
81. Slamon DJ, Godolphin W, Jones LA, et al. Studies of the *HER-2/neu* proto-oncogene in human breast and ovarian cancer. *Science.* 244:707, 1989

82. Slebos RJC, Evers SG, Wagenaar SS, Rodenhuis S. Cellular protooncogenes are infrequently amplified in untreated non-small cell lung cancer. *Br J Cancer.* 59:76, 1989
83. Stern DF, Kamps MP, Cao H. Oncogenic activation of p185^{neu} stimulates tyrosine phosphorylation in vivo. *Mol Cell Biol.* 8:3969, 1988
84. Tal M, Wetzler M, Josefberg Z, et al. Sporadic amplification of the *HER2/neu* protooncogene in adenocarcinomas of various tissues. *Cancer Res.* 48:1517, 1988
85. Tandon AK, Clark GM, Chamness GC, et al. *HER-2/neu* oncogene protein and prognosis in breast cancer. *J Clin Oncol.* 7:1120, 1989
86. Thor AD, Schwartz LH, Koerner FC, et al. Analysis of c-*erb*B-2 expression in breast carcinomas with clinical follow-up. *Cancer Res.* 49:7147, 1989
87. Tsuda H, Hirohashi S, Shimosato Y, et al. Correlation between long-term survival in breast cancer patients and amplification of two putative oncogene-coamplification units: *hst-1/int-2* and c-*erb*B-2/*ear-1*. *Cancer Res.* 49:3104, 1989
88. Tsutsumi Y, Naber SP, DeLellis RA, et al. *Neu* oncogene protein and epidermal growth factor receptor are independently expressed in benign and malignant breast tissues. *Hum Pathol.* 21:750, 1990
89. Tsutsumi Y, Stork PJ, Wolfe HJ. Detection of DNA amplification and mRNA overexpression of the *neu* oncogene in breast carcinomas by polymerase chain reaction (abstract). *Mod Pathol.* 3:101A, 1990
90. Van de Vijver M, van de Bersselaar R, Devilee P, et al. Amplification of the *neu* (c-*erb*B-2) oncogene in human mammary tumors is relatively frequent and is often accompanied by amplification of the linked c-*erb*A oncogene. *Mol Cell Biol.* 7:2019, 1987
91. Van de Vijver MJ, Mooi WJ, Wisman P, et al. Immunohistochemical detection of the *neu* protein in tissue sections of human breast tumors with amplified *neu* DNA. *Oncogene.* 2:175, 1988
92. Van de Vijver MJ, Peterse JL, Mooi WJ, et al. *Neu*-protein overexpression in breast cancer: Association with comedo-type ductal carcinoma in situ and limited prognostic value in stage II breast cancer. *N Engl J Med.* 319:1239, 1988
93. Varley JM, Swallow JE, Brammar WJ, et al. Alterations to either c-*erb*B-2 (*neu*) or c-*myc* proto-oncogenes in breast carcinomas correlate with poor short-term prognosis. *Oncogene.* 1:423, 1987
94. Venter DJ, Tuzi NL, Kumar S, Gullick WJ. Overexpression of the c-*erb*B-2 oncoprotein in human breast carcinomas: Immunohistological assessment correlates with gene amplification. *Lancet.* 2:69, 1987
95. Voravud N, Foster CS, Gilbertson JA, et al. Oncogene expression in cholangiocarcinoma and in normal hepatic development. *Hum Pathol.* 20:1163, 1989
96. Walker RA, Senior PV, Jones JL, et al. An immunohistochemical and in situ hybridization study of *c-myc* and *c-erb*B-2 expression in primary human breast carcinomas. *J Pathol.* 158:97, 1989
97. Ware JL, Maygarden SJ, Koontz WW, Strom SC. Differential reactivity with anti-c-*erb*B-2 antiserum among human malignant and benign prostatic tissue (abstract). *Am Assoc Cancer Res Ann Meeting.* 30:437A, 1989
98. Weiner DB, Liu J, Cohen JA, et al. A point mutation in the *neu* oncogene mimics ligand induction of receptor aggregation. *Nature.* 339:230, 1989
99. Weiner DB, Nordberg J, Robinson R, et al. Expression of the *neu* gene-encoded protein (p185^{neu}) in human non-small cell carcinomas of the lung. *Cancer Res.* 50:421, 1990

100. Wells A. The epidermal growth factor receptor and its ligand. In: Benz C, Liu E, eds. *Oncogenes.* Boston, Kluwer Academic Pub; 1989: 143–168
101. Wright C, Angus B, Nicholson S, et al. Expression of c-*erb*B-2 oncoprotein: A prognostic indicator in human breast cancer. *Cancer Res.* 49:2087, 1989
102. Wu A, Colombero A, Low J, et al. Analysis of expression and mutation of the *erb*B-2 gene in breast carcinoma by the polymerase chain reaction (abstract). *Mod Pathol.* 3:108A, 1990
103. Yamamoto T, Ikawa S, Akiyama T, et al. Similarity of protein encoded by the human c-*erb*B-2 gene to epidermal growth factor receptor. *Nature.* 319:230, 1986
104. Yao M, Shuin T, Misaki H, Kubota Y. Enhanced expression of c-*myc* and epidermal growth factor receptor (C-*erb*B-1) genes in primary human renal cancer. *Cancer Res.* 48:6753, 1988
105. Yarden Y, Weinberg RA. Experimental approaches to hypothetical hormones: Detection of a candidate ligand of the *neu* protooncogene. *Proc Natl Acad Sci USA.* 86:3179, 1989
106. Yee LD, Kacinski BM, Carter D. Oncogene structure, function and expression in breast cancer. *Semin Diagn Pathol.* 6:110, 1989
107. Yokota J, Yamamoto T, Toyoshima K, et al. Amplification of c-*erb*B-2 oncogene in human adenocarcinomas in vivo. *Lancet.* 1:765, 1986
108. Yokota J, Yamamoto T, Miyajima N, et al. Genetic alterations of the c-*erb*B-2 oncogene occur frequently in tubular adenocarcinoma of the stomach and are often accompanied by amplification of the v-*erb*A homologue. *Oncogene.* 2:283, 1988
109. Zeillinger R, Kury F, Czerwenka K, et al. *HER-2* amplification, steroid receptors and epidermal growth factor receptor in primary breast cancer. *Oncogene.* 4:109, 1989
110. Zhang X, Silva E, Gershenson D, Hung MC. Amplification and rearrangement of c-*erb*B proto-oncogenes in cancer of human female genital tract. *Oncogene.* 4:985, 1989
111. Zhou D, Battifora H, Yokota J, et al. Association of multiple copies of the c-*erb*B-2 oncogene with spread of breast cancer. *Cancer Res.* 47:6123, 1987
112. Zhou D, Gonzalez-Cadavid N, Ahuja H, et al. A unique pattern of proto-oncogene abnormalities in ovarian adenocarcinomas. *Cancer.* 62:1573, 1988
113. Zhou D, Ahuja H, Cline MJ. Proto-oncogene abnormalities in human breast cancer: c-*erb*B-2 amplification does not correlate with recurrence of disease. *Oncogene.* 4:105, 1989

8 8

GNE2930R1C3

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

COPY

Applicant	:	Botstein, et al.
Appl. No.	:	10/032,996
Filed	:	December 27, 2001
For	:	SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
Examiner	:	Fredman, J.
Group Art Unit	:	1634

DECLARATION OF PAUL POLAKIS, PH.D. UNDER 37 C.F.R. § 1.132

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

Attached is the Declaration of Paul Polakis, Ph.D.

Respectfully submitted,

KNOBBE, MARTENS, OLSON & BEAR, LLP

Dated: June 16, 2004

By: _____
AnneMarie Kaiser
Registration No. 37,649
Attorney of Record
Customer No. 30,313
(619) 235-8550

S:\DOCS\BSG\BSG-1215.DOC
061404

DECLARATION OF PAUL POLAKIS, Ph.D.

I, Paul Polakis, Ph.D., declare and say as follows:

1. I was awarded a Ph.D. by the Department of Biochemistry of the Michigan State University in 1984. My scientific Curriculum Vitae is attached to and forms part of this Declaration (Exhibit A).

2. I am currently employed by Genentech, Inc. where my job title is Staff Scientist. Since joining Genentech in 1999, one of my primary responsibilities has been leading Genentech's Tumor Antigen Project, which is a large research project with a primary focus on identifying tumor cell markers that find use as targets for both the diagnosis and treatment of cancer in humans.

3. As part of the Tumor Antigen Project, my laboratory has been analyzing differential expression of various genes in tumor cells relative to normal cells. The purpose of this research is to identify proteins that are abundantly expressed on certain tumor cells and that are either (i) not expressed, or (ii) expressed at lower levels, on corresponding normal cells. We call such differentially expressed proteins "tumor antigen proteins". When such a tumor antigen protein is identified, one can produce an antibody that recognizes and binds to that protein. Such an antibody finds use in the diagnosis of human cancer and may ultimately serve as an effective therapeutic in the treatment of human cancer.

4. In the course of the research conducted by Genentech's Tumor Antigen Project, we have employed a variety of scientific techniques for detecting and studying differential gene expression in human tumor cells relative to normal cells, at genomic DNA, mRNA and protein levels. An important example of one such technique is the well known and widely used technique of microarray analysis which has proven to be extremely useful for the identification of mRNA molecules that are differentially expressed in one tissue or cell type relative to another. In the course of our research using microarray analysis, we have identified approximately 200 gene transcripts that are present in human tumor cells at significantly higher levels than in corresponding normal human cells. To date, we have generated antibodies that bind to about 30 of the tumor antigen proteins expressed from these differentially expressed gene transcripts and have used these antibodies to quantitatively determine the level of production of these tumor antigen proteins in both human cancer cells and corresponding normal cells. We have then compared the levels of mRNA and protein in both the tumor and normal cells analyzed.

5. From the mRNA and protein expression analyses described in paragraph 4 above, we have observed that there is a strong correlation between changes in the level of mRNA present in any particular cell type and the level of protein

expressed from that mRNA in that cell type. In approximately 80% of our observations we have found that increases in the level of a particular mRNA correlates with changes in the level of protein expressed from that mRNA when human tumor cells are compared with their corresponding normal cells.

6. Based upon my own experience accumulated in more than 20 years of research, including the data discussed in paragraphs 4 and 5 above and my knowledge of the relevant scientific literature, it is my considered scientific opinion that for human genes, an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell. In fact, it remains a central dogma in molecular biology that increased mRNA levels are predictive of corresponding increased levels of the encoded protein. While there have been published reports of genes for which such a correlation does not exist, it is my opinion that such reports are exceptions to the commonly understood general rule that increased mRNA levels are predictive of corresponding increased levels of the encoded protein.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

Dated: 5/07/04

By: Paul Polakis

Paul Polakis, Ph.D.

SV 2031808 v1

CURRICULUM VITAE

PAUL G. POLAKIS
Staff Scientist
Genentech, Inc
1 DNA Way, MS#40
S. San Francisco, CA 94080

EDUCATION:

Ph.D., Biochemistry, Department of Biochemistry,
Michigan State University (1984)

B.S., Biology. College of Natural Science, Michigan State University (1977)

PROFESSIONAL EXPERIENCE:

2002-present	Staff Scientist, Genentech, Inc S. San Francisco, CA
1999- 2002	Senior Scientist, Genentech, Inc., S. San Francisco, CA
1997 -1999	Research Director Onyx Pharmaceuticals, Richmond, CA
1992- 1996	Senior Scientist, Project Leader, Onyx Pharmaceuticals, Richmond, CA
1991-1992	Senior Scientist, Chiron Corporation, Emeryville, CA.
1989-1991	Scientist, Cetus Corporation, Emeryville CA.
1987-1989	Postdoctoral Research Associate, Genentech, Inc., South SanFrancisco, CA.
1985-1987	Postdoctoral Research Associate, Department of Medicine, Duke University Medical Center, Durham, NC

1984-1985	Assistant Professor, Department of Chemistry, Oberlin College, Oberlin, Ohio
1980-1984	Graduate Research Assistant, Department of Biochemistry, Michigan State University East Lansing, Michigan

PUBLICATIONS:

1. Polakis, P G. and Wilson, J. E. 1982 Purification of a Highly Bindable Rat Brain Hexokinase by High Performance Liquid Chromatography. **Biochem. Biophys. Res. Commun.** 107, 937-943.

2. Polakis, P.G. and Wilson, J. E. 1984 Proteolytic Dissection of Rat Brain Hexokinase: Determination of the Cleavage Pattern during Limited Digestion with Trypsin. **Arch. Biochem. Biophys.** 234, 341-352.

3. Polakis, P. G. and Wilson, J. E. 1985 An Intact Hydrophobic N-Terminal Sequence is Required for the Binding Rat Brain Hexokinase to Mitochondria. **Arch. Biochem. Biophys.** 236, 328-337.

4. Uhing, R.J., **Polakis,P.G.** and Snyderman, R. 1987 Isolaton of GTP-binding Proteins from Myeloid HL60 Cells. **J. Biol. Chem.** 262, 15575-15579.

5. Polakis, P.G., Uhing, R.J. and Snyderman, R. 1988 The Formylpeptide Chemoattractant Receptor Copurifies with a GTP-binding Protein Containing a Distinct 40 kDa Pertussis Toxin Substrate. **J. Biol. Chem.** 263, 4969-4979.

6. Uhing, R. J., Dillon, S., **Polakis, P. G.**, Truett, A. P. and Snyderman, R. 1988 Chemoattractant Receptors and Signal Transduction Processes in Cellular and Molecular Aspects of Inflammation (Poste, G. and Crooke, S. T. eds.) pp 335-379.

7. Polakis, P.G., Evans, T. and Snyderman 1989 Multiple Chromatographic Forms of the Formylpeptide Chemoattractant Receptor and their Relationship to GTP-binding Proteins. **Biochem. Biophys. Res. Commun.** 161, 276-283.

8. Polakis, P. G., Snyderman, R. and Evans, T. 1989 Characterization of G25K, a GTP-binding Protein Containing a Novel Putative Nucleotide Binding Domain. **Biochem. Biophys. Res. Comun.** 160, 25-32.

9. Polakis, P., Weber,R.F., Nevins,B., Didsbury, J. Evans,T. and Snyderman, R. 1989 Identification of the ral and rac1 Gene Products, Low Molecular Mass GTP-binding Proteins from Human Platelets. **J. Biol. Chem.** 264, 16383-16389.

10. Snyderman, R., Perianin, A., Evans, T., **Polakis, P.** and Didsbury, J. 1989 G Proteins and Neutrophil Function. In ADP-Ribosylating Toxins and G Proteins: Insights into Signal Transduction. (J. Moss and M. Vaughn, eds.) Amer. Soc. Microbiol. pp. 295-323.

11. Hart, M.J., **Polakis, P.G.**, Evans, T. and Cerrione, R.A. 1990 The Identification and Charaterization of an Epidermal Growth Factor-Stimulated Phosphorylation of a Specific Low Molecular Mass GTP-binding Protein in a Reconstituted Phospholipid Vesicle System. **J. Biol. Chem.** 265, 5990-6001.

12. Yatani, A., Okabe, K., **Polakis, P.** Halenbeck, R. McCormick, F. and Brown, A. M. 1990 ras p21 and GAP Inhibit Coupling of Muscarinic Receptors to Atrial K^+ Channels. **Cell.** 61, 769-776.

13. Munemitsu, S., Innis, M.A., Clark, R., McCormick, F., Ullrich, A. and **Polakis, P.G.** 1990 Molecular Cloning and Expression of a G25K cDNA, the Human Homolog of the Yeast Cell Cycle Gene CDC42. **Mol. Cell. Biol.** 10, 5977-5982.

14. Polakis, P.G. Rubinfeld, B. Evans, T. and McCormick, F. 1991 Purification of Plasma Membrane-Associated GTPase Activating Protein (GAP) Specific for rap-1/krev-1 from HL60 Cells. **Proc. Natl. Acad. Sci. USA** 88, 239-243.

15. Moran, M. F., **Polakis, P.**, McCormick, F., Pawson, T. and Ellis, C. 1991 Protein Tyrosine Kinases Regulate the Phosphorylation, Protein Interactions, Subcellular Distribution, and Activity of p21ras GTPase Activating Protein. **Mol. Cell. Biol.** 11, 1804-1812

16. Rubinfeld, B., Wong, G., Bekesi, E. Wood, A. McCormick, F. and **Polakis, P. G.** 1991 A Synthetic Peptide Corresponding to a Sequence in the GTPase Activating Protein Inhibits $p21^{ras}$ Stimulation and Promotes Guanine Nucleotide Exchange. **Internatl. J. Peptide and Prot. Res.** 38, 47-53.

17. Rubinfeld, B., Munemitsu, S., Clark, R., Conroy, L., Watt, K., Crosier, W., McCormick, F., and **Polakis, P.** 1991 Molecular Cloning of a GTPase Activating Protein Specific for the Krev-1 Protein $p21^{rap1}$. **Cell** 65, 1033-1042.

18. Zhang, K. Papageorge, A., G., Martin, P., Vass, W. C., Olah, Z., **Polakis, P.**, McCormick, F. and Lowy, D, R. 1991 Heterogenous Amino Acids in RAS and Rap1A Specifying Sensitivity to GAP Proteins. **Science** 254, 1630-1634.

19. Martin, G., Yatani, A., Clark, R., **Polakis, P.**, Brown, A. M. and McCormick, F. 1992 GAP Domains Responsible for $p21^{ras}$-dependent Inhibition of Muscarinic Atrial K^+ Channel Currents. **Science** 255, 192-194.

20. McCormick, F., Martin, G. A., Clark, R., Bollag, G. and **Polakis, P** . 1992 Regulation of p21ras by GTPase Activating Proteins. Cold Spring Harbor **Symposia on Quantitative Biology**. Vol. 56, 237-241.

21. Pronk, G. B., **Polakis, P.**, Wong, G., deVries-Smits, A. M., Bos J. L. and McCormick, F. 1992 $p60^{v\text{-}src}$ Can Associate with and Phosphorylate the $p21^{ras}$ GTPase Activating Protein. **Oncogene** 7,389-394.

22. Polakis P. and McCormick, F. 1992 Interactions Between $p21^{ras}$ Proteins and Their GTPase Activating Proteins. In **Cancer Surveys** (Franks, L. M., ed.) 12, 25-42.

23. Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M., **Polakis,P.** and McCormick, F. 1992 Molecular coloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. **Cell** 69, 551-558.

24. Polakis, P., Rubinfeld, B. and McCormick, F. 1992 Phosphorylation of rap1GAP in vivo and by cAMP-dependent Kinase and the Cell Cycle $p34^{cdc2}$ Kinase in vitro. **J. Biol. Chem.** 267, 10780-10785.

25. McCabe, P.C., Haubrauck, H., **Polakis, P.**, McCormick, F., and Innis, M. A. 1992 Functional Interactions Between $p21^{rap1A}$ and Components of the Budding pathway of *Saccharomyces cerevisiae*. **Mol. Cell. Biol.** 12, 4084-4092.

26. Rubinfeld, B., Crosier, W.J., Albert, I., Conroy,L., Clark, R., McCormick, F. and **Polakis, P.** 1992 Localization of the rap1GAP Catalytic Domain and Sites of Phosphorylation by Mutational Analysis. **Mol. Cell . Biol.** 12, 4634-4642.

27. Ando, S., Kaibuchi, K., Sasaki, K., Hiraoka, T., Nishiyama, T., Mizuno, T., Asada, M., Nunoi, H., Matsuda, I., Matsuura, Y., **Polakis, P.**, McCormick, F. and Takai, Y. 1992 Post-translational processing of rac p21s is important both for their interaction with the GDP/GTP exchange proteins and for their activation of NADPH oxidase. **J. Biol. Chem.** 267, 25709-25713.

28. Janoueix-Lerosey, I., **Polakis, P.**, Tavitian, A. and deGunzberg, J. 1992 Regulation of the GTPase activity of the ras-related rap2 protein. **Biochem. Biophys. Res. Commun.** 189, 455-464.

29. Polakis, P. 1993 GAPs Specific for the rap1/Krev-1 Protein. in GTP-binding Proteins: the *ras*-superfamily. (J.C. LaCale and F. McCormick, eds.) 445-452.

30. Polakis, P. and McCormick, F. 1993 Structural requirements for the interaction of $p21^{ras}$ with GAP, exchange factors, and its bological effector target. **J. Biol Chem.** 268, 9157-9160.

31. Rubinfeld, B., Souza, B. Albert, I., Muller, O., Chamberlain, S., Masiarz, F., Munemitsu, S. and **Polakis, P.** 1993 Association of the APC gene product with beta- catenin. **Science** 262, 1731-1734.

32. Weiss, J., Rubinfeld, B., **Polakis, P.**, McCormick, F. Cavenee, W. A. and Arden, K. 1993 The gene for human rap1-GTPase activating protein (rap1GAP) maps to chromosome 1p35-1p36.1. **Cytogenet. Cell Genet.** 66, 18-21.

33. Sato, K. Y., **Polakis, P.**, Haubruck, H., Fasching, C. L., McCormick, F. and Stanbridge, E. J. 1994 Analysis of the tumor suppressor acitvity of the K-rev gene in human tumor cell lines. **Cancer Res.** 54, 552-559.

34. Janoueix-Lerosey, I., Fontenay, M., Tobelem, G., Tavitian, A., **Polakis, P.** and DeGunzburg, J. 1994 Phosphorylation of rap1GAP during the cell cycle. **Biochem. Biophys. Res. Commun.** 202, 967-975

35. Munemitsu, S., Souza, B., Mueller, O., Albert, I., Rubinfeld, B., and **Polakis, P.** 1994 The APC gene product associates with microtubules in vivo and affects their assembly in vitro. **Cancer Res.** 54, 3676-3681.

36. Rubinfeld, B. and **Polakis, P.** 1995 Purification of baculovirus produced rap1GAP. **Methods Enz.** 255,31

37. **Polakis, P.** 1995 Mutations in the APC gene and their implications for protein structure and function. **Current Opinions in Genetics and Development** 5, 66-71

38. Rubinfeld, B., Souza, B., Albert, I., Munemitsu, S. and **Polakis P.** 1995 The APC protein and E-cadherin form similar but independent complexes with α-catenin, β-catenin and Plakoglobin. **J. Biol. Chem.** 270, 5549-5555

39. Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B., and **Polakis, P.** 1995 Regulation of intracellular β-catenin levels by the APC tumor suppressor gene. **Proc. Natl. Acad. Sci.** 92, 3046-3050.

40. Lock, P., Fumagalli, S., **Polakis, P.** McCormick, F. and Courtneidge, S. A. 1996 The human p62 cDNA encodes Sam68 and not the rasGAP-associated p62 protein. **Cell** 84, 23-24.

41. Papkoff, J., Rubinfeld, B., Schryver, B. and **Polakis, P.** 1996 Wnt-1 regulates free pools of catenins and stabilizes APC-catenin complexes. **Mol. Cell. Biol.** 16, 2128-2134.

42. Rubinfeld, B., Albert, I., Porfiri, E., Fiol, C., Munemitsu, S. and **Polakis, P.** 1996 Binding of GSK3β to the APC-β-catenin complex and regulation of complex assembly. **Science** 272, 1023-1026.

43. Munemitsu, S., Albert, I., Rubinfeld, B. and **Polakis, P.** 1996 Deletion of amino-terminal structure stabilizes β-catenin in vivo and promotes the hyperphosphorylation of the APC tumor suppressor protein. **Mol. Cell. Biol.**16, 4088-4094.

44. Hart, M. J., Callow, M. G., Sousa, B. and **Polakis P.** 1996 IQGAP1, a calmodulin binding protein with a rasGAP related domain, is a potential effector for cdc42Hs. **EMBO J.** 15, 2997-3005.

45. Nathke, I. S., Adams, C. L., **Polakis, P.**, Sellin, J. and Nelson, W. J. 1996 The adenomatous polyposis coli (APC) tumor suppressor protein is localized to plasma membrane sites involved in active epithelial cell migration. **J. Cell. Biol.** 134, 165-180.

46. Hart, M. J., Sharma, S., elMasry, N., Qui, R-G., McCabe, P., **Polakis, P.** and Bollag, G. 1996 Identification of a novel guanine nucleotide exchange factor for the rho GTPase. **J. Biol. Chem.** 271, 25452.

47. Thomas JE, Smith M, Rubinfeld B, Gutowski M, Beckmann RP, and **Polakis P.** 1996 Subcellular localization and analysis of apparent 180-kDa and 220-kDa proteins of the breast cancer susceptibility gene, BRCA1. **J. Biol. Chem.** 1996 271, 28630-28635

48. Hayashi, S., Rubinfeld, B., Souza, B., **Polakis, P.**, Wieschaus, E., and Levine, A. 1997 A Drosophila homolog of the tumor suppressor adenomatous polyposis coli

down-regulates β -catenin but its zygotic expression is not essential for the regulation of armadillo. **Proc. Natl. Acad. Sci.** 94, 242-247.

49. Vleminckx, K., Rubinfeld, B., **Polakis, P.** and Gumbiner, B. 1997 The APC tumor suppressor protein induces a new axis in Xenopus embryos. **J. Cell. Biol.** 136, 411-420.

50. Rubinfeld, B., Robbins, P., El-Gamil, M., Albert, I., Porfiri, P. and **Polakis, P.** 1997 Stabilization of β-catenin by genetic defects in melanoma cell lines. **Science** 275, 1790-1792.

51. **Polakis, P.** The adenomatous polyposis coli (APC) tumor suppressor. 1997 **Biochem. Biophys. Acta**, 1332, F127-F147.

52. Rubinfeld, B., Albert, I., Porfiri, E., Munemitsu, S., and **Polakis, P** 1997 Loss of β-catenin regulation by the APC tumor suppressor protein correlates with loss of structure due to common somatic mutations of the gene. **Cancer Res.** 57, 4624-4630.

53. Porfiri, E., Rubinfeld, B., Albert, I., Hovanes. K., Waterman, M., and **Polakis, P.** 1997 Induction of a β-catenin-LEF-1 complex by wnt-1 and transforming mutants of β-catenin. **Oncogene** 15, 2833-2839.

54. Thomas JE, Smith M, Tonkinson JL, Rubinfeld B, and **Polakis P.**, 1997 Induction of phosphorylation on BRCA1 during the cell cycle and after DNA damage. **Cell Growth Differ**. 8, 801-809.

55. Hart, M., de los Santos, R., Albert, I., Rubinfeld, B., and **Polakis P.**, 1998 Down regulation of β-catenin by human Axin and its association with the adenomatous polyposis coli (APC) tumor suppressor, β-catenin and glycogen synthase kinase 3β. **Current Biology 8**, 573-581.

56. Polakis, P. 1998 The oncogenic activation of β-catenin. **Current Opinions in Genetics and Development** 9, 15-21

57. Matt Hart, Jean-Paul Concordet, Irina Lassot, Iris Albert, Rico del los Santos, Herve Durand, Christine Perret, Bonnee Rubinfled, Florence Margottin, Richard Benarous and **Paul Polakis**. 1999 The F-box protein β–TrCP associates with phosphorylated β–catenin and regulates its activity in the cell. **Current Biology** 9, 207-10.

58. Howard C. Crawford, Barbara M. Fingleton, Bonnee Rubinfeld, **Paul Polakis** and Lynn M. Matrisian 1999 The metalloproteinase matrilysin is a target of β–catenin transactivation in intestinal tumours. **Oncogene** 18, 2883-91.

59. Meng J, Glick JL, **Polakis P**, Casey PJ. 1999 Functional interaction between Galpha(z) and Rap1GAP suggests a novel form of cellular cross-talk. **J Biol Chem.** 17, 36663-9

60. Vijayasurian Easwaran, Virginia Song, **Paul Polakis** and Steve Byers 1999 The ubiquitin-proteosome pathway and serine kinase activity modulate APC mediated regulation of β-catenin-LEF signaling. **J. Biol. Chem.** 274(23):16641-5.

61 Polakis P, Hart M and Rubinfeld B. 1999 Defects in the regulation of beta-catenin in colorectal cancer. **Adv Exp Med Biol.** 470, 23-32

62 Shen Z, Batzer A, Koehler JA, **Polakis P**, Schlessinger J, Lydon NB, Moran MF. 1999 Evidence for SH3 domain directed binding and phosphorylation of Sam68 by Src. **Oncogene.** 18, 4647-53

64. Thomas GM, Frame S, Goedert M, Nathke I, **Polakis P**, Cohen P. 1999 A GSK3- binding peptide from FRAT1 selectively inhibits the GSK3-catalysed phosphorylation of axin and beta-catenin. **FEBS Lett.** 458, 247-51.

65. Peifer M, **Polakis P.** 2000 Wnt signaling in oncogenesis and embryogenesis--a look outside the nucleus. **Science** 287,1606-9,

66. **Polakis P.** 2000 Wnt signaling and cancer. **Genes Dev**;14, 1837-1851.

67. Spink KE, **Polakis P**, Weis WI 2000 Structural basis of the Axin-adenomatous polyposis coli interaction. **EMBO J** 19, 2270-2279.

68. Szeto , W., Jiang, W., Tice, D.A., Rubinfeld, B., Hollingshead, P.G., Fong, S.E., Dugger, D.L., Pham, T., Yansura, D.E., Wong, T.A., Grimaldi, J.C., Corpuz, R.T., Singh J.S., Frantz, G.D., Devaux, B., Crowley, C.W., Schwall, R.H., Eberhard, D.A., Rastelli, L., **Polakis, P.** and Pennica, D. 2001 Overexpression of the Retinoic Acid-Responsive Gene Stra6 in Human Cancers and its Synergistic Induction by Wnt-1 and Retinoic Acid. **Cancer Res** 61, 4197-4204.

69. Rubinfeld B, Tice DA, **Polakis P.** 2001 Axin dependent phosphorylation of the adenomatous polyposis coli protein mediated by casein kinase 1 epsilon. **J Biol Chem** 276, 39037-39045.

70. **Polakis P.** 2001 More than one way to skin a catenin. **Cell** 2001 105, 563-566.

71. Tice DA, Soloviev I, **Polakis P.** 2002 Activation of the Wnt Pathway Interferes withSerum Response Element-driven Transcription of Immediate Early Genes. **J Biol. Chem.** 277, 6118-6123.

72. Tice DA, Szeto W, Soloviev I, Rubinfeld B, Fong SE, Dugger DL, Winer J,

Williams PM, Wieand D, Smith V, Schwall RH, Pennnica D, **Polakis P.** 2002 Synergistic activation of tumor antigens by wnt-1 signaling and retinoic acid revealed by gene expression profiling. **J Biol Chem.** 277,14329-14335.

73. **Polakis, P.** 2002 Casein kinase I: A wnt'er of disconnect. **Curr. Biol.** 12, R499.

74. Mao,W. , Luis, E., Ross, S., Silva, J., Tan, C., Crowley, C., Chui, C., Franz, G., Senter, P., Koeppen, H., **Polakis, P.** 2004 EphB2 as a therapeutic antibody drug target for the treatment of colorectal cancer. **Cancer Res.** 64, 781-788.

75. Shibamoto, S., Winer, J., Williams, M., Polakis, P. 2003 A Blockade in Wnt signaling is activated following the differentiation of F9 teratocarcinoma cells. **Exp. Cell Res.** 29211-20.

76. Zhang Y, Eberhard DA, Frantz GD, Dowd P, Wu TD, Zhou Y, Watanabe C, Luoh SM, **Polakis P,** Hillan KJ, Wood WI, Zhang Z. 2004 GEPIS--quantitative gene expression profiling in normal and cancer tissues. **Bioinformatics**, April 8

MOLECULAR BIOLOGY OF
THE CELL
THIRD EDITION

Text Editor: Miranda Robertson
Managing Editor: Ruth Adams
Illustrator: Nigel Orme
Molecular Model Drawings: Kate Hesketh-Moore
Director of Electronic Publishing: John M-Roblin
Computer Specialist: Chuck Bartelt
Disk Preparation: Carol Winter
Copy Editor: Shirley M. Cobert
Production Editor: Douglas Goertzen
Production Coordinator: Perry Bessas
Indexer: Maija Hinkle

Bruce Alberts received his Ph.D. from Harvard University and is currently President of the National Academy of Sciences and Professor of Biochemistry and Biophysics at the University of California, San Francisco. *Dennis Bray* received his Ph.D. from the Massachusetts Institute of Technology and is currently a Medical Research Council Fellow in the Department of Zoology, University of Cambridge. *Julian Lewis* received his D.Phil. from the University of Oxford and is currently a Senior Scientist in the Imperial Cancer Research Fund Developmental Biology Unit, University of Oxford. *Martin Raff* received his M.D. from McGill University and is currently a Professor in the MRC Laboratory for Molecular Cell Biology and the Biology Department, University College London. *Keith Roberts* received his Ph.D. from the University of Cambridge and is currently Head of the Department of Cell Biology, the John Innes Institute, Norwich. *James D. Watson* received his Ph.D. from Indiana University and is currently Director of the Cold Spring Harbor Laboratory. He is the author of *Molecular Biology of the Gene* and, with Francis Crick and Maurice Wilkins, won the Nobel Prize in Medicine and Physiology in 1962.

Library of Congress Cataloging-in-Publication Data
Molecular biology of the cell / Bruce Alberts . . . [et al.].—3rd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-8153-1619-4 (hard cover).—ISBN 0-8153-1620-8 (pbk.)
1. Cytology. 2. Molecular biology. I. Alberts, Bruce.
[DNLM: 1. Cells. 2. Molecular Biology. QH 581.2 M718 1994]
QH581.2.M64 1994
574.87—dc20
DNLM/DLC
for Library of Congress 93-45907
CIP

Published by Garland Publishing, Inc.
717 Fifth Avenue, New York, NY 10022

Printed in the United States of America

15 14 13 12 10 9 8 7

Front cover: The photograph shows a rat nerve cell in culture. It is labeled (*yellow*) with a fluorescent antibody that stains its cell body and dendritic processes. Nerve terminals (*green*) from other neurons (not visible), which have made synapses on the cell, are labeled with a different antibody. (Courtesy of Olaf Mundigl and Pietro de Camilli.)

Dedication page: Gavin Borden, late president of Garland Publishing, weathered in during his mid-1980s climb near Mount McKinley with MBoC author Bruce Alberts and famous mountaineer guide Mugs Stump (1940–1992).

Back cover: The authors, in alphabetical order, crossing Abbey Road in London on their way to lunch. Much of this third edition was written in a house just around the corner. (Photograph by Richard Olivier.)

extracts. If these minor cell proteins differ among cells to the same extent as the more abundant proteins, as is commonly assumed, only a small number of protein differences (perhaps several hundred) suffice to create very large differences in cell morphology and behavior.

A Cell Can Change the Expression of Its Genes in Response to External Signals [3]

Most of the specialized cells in a multicellular organism are capable of altering their patterns of gene expression in response to extracellular cues. If a liver cell is exposed to a glucocorticoid hormone, for example, the production of several specific proteins is dramatically increased. Glucocorticoids are released during periods of starvation or intense exercise and signal the liver to increase the production of glucose from amino acids and other small molecules; the set of proteins whose production is induced includes enzymes such as tyrosine aminotransferase, which helps to convert tyrosine to glucose. When the hormone is no longer present, the production of these proteins drops to its normal level.

Other cell types respond to glucocorticoids in different ways. In fat cells, for example, the production of tyrosine aminotransferase is reduced, while some other cell types do not respond to glucocorticoids at all. These examples illustrate a general feature of cell specialization—different cell types often respond in different ways to the same extracellular signal. Underlying this specialization are features that do not change, which give each cell type its permanently distinctive character. These features reflect the persistent expression of different sets of genes.

Gene Expression Can Be Regulated at Many of the Steps in the Pathway from DNA to RNA to Protein [4]

If differences between the various cell types of an organism depend on the particular genes that the cells express, at what level is the control of gene expression exercised? There are many steps in the pathway leading from DNA to protein, and all of them can in principle be regulated. Thus a cell can control the proteins it makes by (1) controlling when and how often a given gene is transcribed (**transcriptional control**), (2) controlling how the primary RNA transcript is spliced or otherwise processed (**RNA processing control**), (3) selecting which completed mRNAs in the cell nucleus are exported to the cytoplasm (**RNA transport control**), (4) selecting which mRNAs in the cytoplasm are translated by ribosomes (**translational control**), (5) selectively destabilizing certain mRNA molecules in the cytoplasm (**mRNA degradation control**), or (6) selectively activating, inactivating, or compartmentalizing specific protein molecules after they have been made (**protein activity control**) (Figure 9–2).

For most genes transcriptional controls are paramount. This makes sense because, of all the possible control points illustrated in Figure 9–2, only transcriptional control ensures that no superfluous intermediates are synthesized. In the

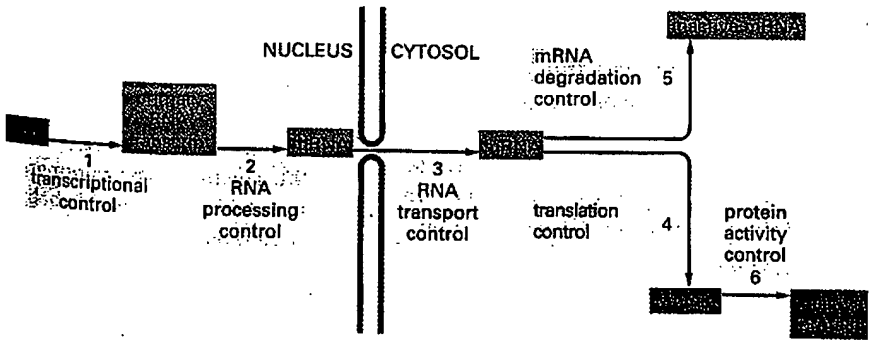


Figure 9–2 Six steps at which eucaryote gene expression can be controlled. Only controls that operate at steps 1 through 5 are discussed in this chapter. The regulation of protein activity (step 6) is discussed in Chapter 5; this includes reversible activation or inactivation by protein phosphorylation as well as irreversible inactivation by proteolytic degradation.

following sections we discuss the DNA and protein components that regulate the initiation of gene transcription. We return at the end of the chapter to the other ways of regulating gene expression.

Summary

The genome of a cell contains in its DNA sequence the information to make many thousands of different protein and RNA molecules. A cell typically expresses only a fraction of its genes, and the different types of cells in multicellular organisms arise because different sets of genes are expressed. Moreover, cells can change the pattern of genes they express in response to changes in their environment, such as signals from other cells. Although all of the steps involved in expressing a gene can in principle be regulated, for most genes the initiation of RNA transcription is the most important point of control.

DNA-binding Motifs in Gene Regulatory Proteins [5]

How does a cell determine which of its thousands of genes to transcribe? As discussed in Chapter 8, the transcription of each gene is controlled by a regulatory region of DNA near the site where transcription begins. Some regulatory regions are simple and act as switches that are thrown by a single signal. Other regulatory regions are complex and act as tiny microprocessors, responding to a variety of signals that they interpret and integrate to switch the neighboring gene on or off. Whether complex or simple, these switching devices consist of two fundamental types of components: (1) short stretches of DNA of defined sequence and (2) *gene regulatory proteins* that recognize and bind to them.

We begin our discussion of gene regulatory proteins by describing how these proteins were discovered.

Gene Regulatory Proteins Were Discovered Using Bacterial Genetics [6]

Genetic analyses in bacteria carried out in the 1950s provided the first evidence of the existence of **gene regulatory proteins** that turn specific sets of genes on or off. One of these regulators, the *lambda repressor,* is encoded by a bacterial virus, *bacteriophage lambda.* The repressor shuts off the viral genes that code for the protein components of new virus particles and thereby enables the viral genome to remain a silent passenger in the bacterial chromosome, multiplying with the bacterium when conditions are favorable for bacterial growth (see Figure 6–80). The lambda repressor was among the first gene regulatory proteins to be characterized, and it remains one of the best understood, as we discuss later. Other bacterial regulators respond to nutritional conditions by shutting off genes encoding specific sets of metabolic enzymes when they are not needed. The *lac repressor,* for example, the first of these bacterial proteins to be recognized, turns off the production of the proteins responsible for lactose metabolism when this sugar is absent from the medium.

The first step toward understanding gene regulation was the isolation of mutant strains of bacteria and bacteriophage lambda that were unable to shut off specific sets of genes. It was proposed at the time, and later proved, that most of these mutants were deficient in proteins acting as specific repressors for these sets of genes. Because these proteins, like most gene regulatory proteins, are present in small quantities, it was difficult and time-consuming to isolate them. They were eventually purified by fractionating cell extracts on a series of standard chromatography columns (see pp. 166–169). Once isolated, the proteins were shown to bind to specific DNA sequences close to the genes that they

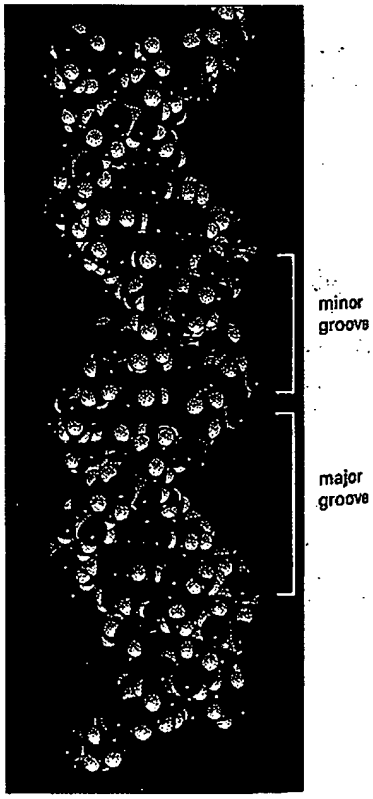


Figure 9–3 Double-helical structure of DNA. The major and minor grooves on the outside of the double helix are indicated. The atoms are colored as follows: carbon, *dark blue;* nitrogen, *light blue;* hydrogen, *white;* oxygen, *red;* phosphorus, *yellow.*

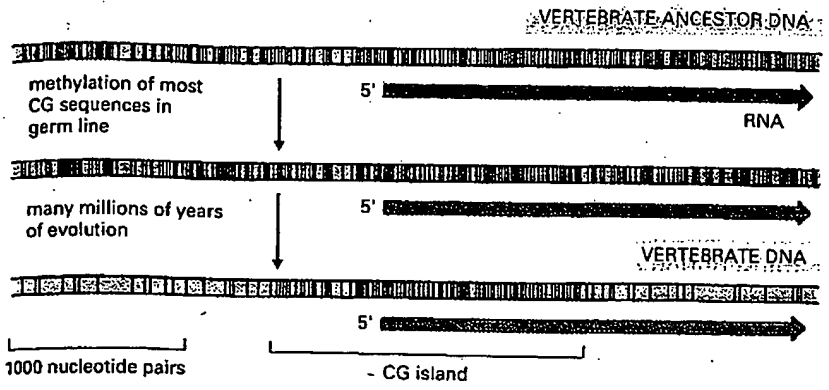


Figure 9–71 A mechanism to explain both the marked deficiency of CG sequences and the presence of CG islands in vertebrate genomes. A *black line* marks the location of an unmethylated CG dinucleotide in the DNA sequence, while a *red line* marks the location of a methylated CG dinucleotide.

Summary

The many types of cells in animals and plants are created largely through mechanisms that cause different genes to be transcribed in different cells. Since many specialized animal cells can maintain their unique character when grown in culture, the gene regulatory mechanisms involved in creating them must be stable once established and heritable when the cell divides, endowing the cell with a memory of its developmental history. Procaryotes and yeasts provide unusually accessible model systems in which to study gene regulatory mechanisms, some of which may be relevant to the creation of specialized cell types in higher eucaryotes. One such mechanism involves a competitive interaction between two (or more) gene regulatory proteins, each of which inhibits the synthesis of the other; this can create a flip-flop switch that switches a cell between two alternative patterns of gene expression. Direct or indirect positive feedback loops, which enable gene regulatory proteins to perpetuate their own synthesis, provide a general mechanism for cell memory.

In eucaryotes gene transcription is generally controlled by combinations of gene regulatory proteins. It is thought that each type of cell in a higher eucaryotic organism contains a specific combination of gene regulatory proteins that ensures the expression of only those genes appropriate to that type of cell. A given gene regulatory protein may be expressed in a variety of circumstances and typically is involved in the regulation of many genes.

In addition to diffusible gene regulatory proteins, inherited states of chromatin condensation are also utilized by eucaryotic cells to regulate gene expression. In vertebrates DNA methylation also plays a part, mainly as a device to reinforce decisions about gene expression that are made initially by other mechanisms.

Posttranscriptional Controls

Although controls on the initiation of gene transcription are the predominant form of regulation for most genes, other controls can act later in the pathway from RNA to protein to modulate the amount of gene product that is made. Although these **posttranscriptional controls,** which operate after RNA polymerase has bound to the gene's promoter and begun RNA synthesis, are less common than *transcriptional control,* for many genes they are crucial. It seems that every step in gene expression that could be controlled in principle is likely to be regulated under some circumstances for some genes.

We consider the varieties of posttranscriptional regulation in temporal order, according to the sequence of events that might be experienced by an RNA molecule after its transcription has begun (Figure 9–72).

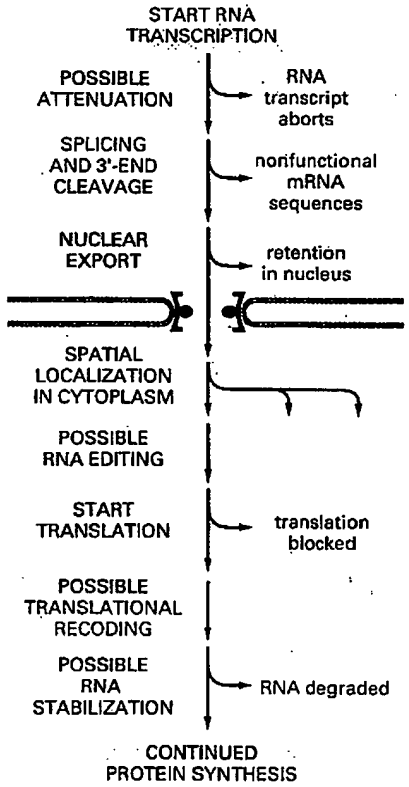


Figure 9–72 Possible posttranscriptional controls on gene expression. Only a few of these controls are likely to be used for any one gene.

MOLECULAR BIOLOGY OF THE CELL

fourth edition

Bruce Alberts

Alexander Johnson

Julian Lewis

Martin Raff

Keith Roberts

Peter Walter

GS Garland Science
Taylor & Francis Group

Garland
Vice President: Denise Schanck
Managing Editor: Sarah Gibbs
Senior Editorial Assistant: Kirsten Jenner
Managing Production Editor: Emma Hunt
Proofreader and Layout: Emma Hunt
Production Assistant: Angela Bennett
Text Editors: Marjorie Singer Anderson and Betsy Dilernia
Copy Editor: Bruce Goatly
Word Processors: Fran Dependahl, Misty Landers and Carol Winter
Designer: Blink Studio, London
Illustrator: Nigel Orme
Indexer: Janine Ross and Sherry Granum
Manufacturing: Nigel Eyre and Marion Morrow

Cell Biology Interactive
Artistic and Scientific Direction: Peter Walter
Narrated by: Julie Theriot
Production, Design, and Development: Mike Morales

Bruce Alberts received his Ph.D. from Harvard University and is President of the National Academy of Sciences and Professor of Biochemistry and Biophysics at the University of California, San Francisco. **Alexander Johnson** received his Ph.D. from Harvard University and is a Professor of Microbiology and Immunology at the University of California, San Francisco. **Julian Lewis** received his D.Phil. from the University of Oxford and is a Principal Scientist at the Imperial Cancer Research Fund, London. **Martin Raff** received his M.D. from McGill University and is at the Medical Research Council Laboratory for Molecular Cell Biology and Cell Biology Unit and in the Biology Department at University College London. **Keith Roberts** received his Ph.D. from the University of Cambridge and is Associate Research Director at the John Innes Centre, Norwich. **Peter Walter** received his Ph.D. from The Rockefeller University in New York and is Professor and Chairman of the Department of Biochemistry and Biophysics at the University of California, San Francisco, and an Investigator of the Howard Hughes Medical Institute.

Library of Congress Cataloging-in-Publicaton Data
Molecular biology of the cell / Bruce Alberts ... [et al.].-- 4th ed.
p. cm
Includes bibliographical references and index.
ISBN 0-8153-3218-1 (hardbound) -- ISBN 0-8153-4072-9 (pbk.)
1. Cytology. 2. Molecular biology. I. Alberts, Bruce.
[DNLM: 1. Cells. 2. Molecular Biology.]
QH581.2 .M64 2002
571.6--dc21

2001054471 CIP

Published by Garland Science, a member of the Taylor & Francis Group, 29 West 35th Street, New York, NY 10001-2299

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

Front cover Human Genome: Reprinted by permission from *Nature*, International Human Genome Sequencing Consortium, 409:860–921, 2001 © Macmillan Magazines Ltd. Adapted from an image by Francis Collins, NHGRI; Jim Kent, UCSC; Ewan Birney, EBI; and Darryl Leja, NHGRI; showing a portion of Chromosome 1 from the initial sequencing of the human genome.

Back cover In 1967, the British artist Peter Blake created a design classic. Nearly 35 years later Nigel Orme (illustrator), Richard Denyer (photographer), and the authors have together produced an affectionate tribute to Mr Blake's image. With its gallery of icons and influences, its assembly created almost as much complexity, intrigue and mystery as the original. *Drosophila, Arabidopsis,* Dolly and the assembled company tempt you to dip inside where, as in the original, "a splendid time is guaranteed for all." (Gunter Blobel, courtesy of The Rockefeller University; Marie Curie, Keystone Press Agency Inc; Darwin bust, by permission of the President and Council of the Royal Society; Rosalind Franklin, courtesy of Cold Spring Harbor Laboratory Archives; Dorothy Hodgkin, © The Nobel Foundation, 1964; James Joyce, etching by Peter Blake; Robert Johnson, photo booth self-portrait early 1930s, © 1986 Delta Haze Corporation all rights reserved, used by permission; Albert L. Lehninger, (unidentified photographer) courtesy of The Alan Mason Chesney Medical Archives of The Johns Hopkins Medical Institutions; Linus Pauling, from Ava Helen and Linus Pauling Papers, Special Collections, Oregon State University; Nicholas Poussin, courtesy of ArtToday.com; Barbara McClintock, © David Micklos, 1983; Andrei Sakharov, courtesy of Elena Bonner; Frederick Sanger, © The Nobel Foundation, 1958.)

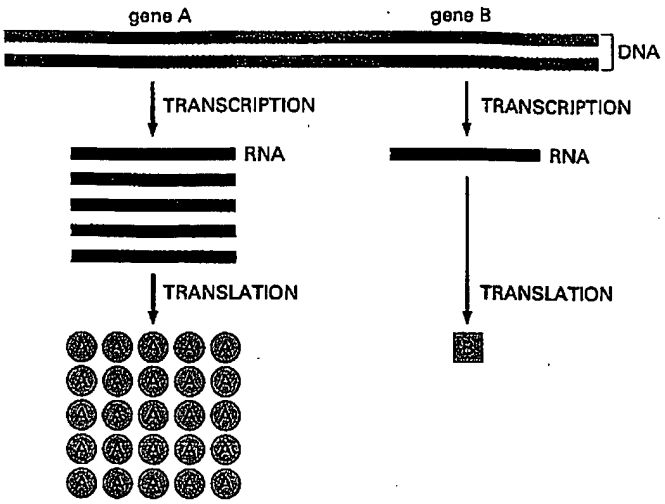


Figure 6–3 Genes can be expressed with different efficiencies. Gene A is transcribed and translated much more efficiently than gene B. This allows the amount of protein A in the cell to be much greater than that of protein B.

FROM DNA TO RNA

Transcription and translation are the means by which cells read out, or express, the genetic instructions in their genes. Because many identical RNA copies can be made from the same gene, and each RNA molecule can direct the synthesis of many identical protein molecules, cells can synthesize a large amount of protein rapidly when necessary. But each gene can also be transcribed and translated with a different efficiency, allowing the cell to make vast quantities of some proteins and tiny quantities of others (Figure 6–3). Moreover, as we see in the next chapter, a cell can change (or regulate) the expression of each of its genes according to the needs of the moment—most obviously by controlling the production of its RNA.

Portions of DNA Sequence Are Transcribed into RNA

The first step a cell takes in reading out a needed part of its genetic instructions is to copy a particular portion of its DNA nucleotide sequence—a gene—into an RNA nucleotide sequence. The information in RNA, although copied into another chemical form, is still written in essentially the same language as it is in DNA—the language of a nucleotide sequence. Hence the name **transcription.**

Like DNA, RNA is a linear polymer made of four different types of nucleotide subunits linked together by phosphodiester bonds (Figure 6–4). It differs from DNA chemically in two respects: (1) the nucleotides in RNA are *ribonucleotides*—that is, they contain the sugar ribose (hence the name *ribo*nucleic acid) rather than deoxyribose; (2) although, like DNA, RNA contains the bases adenine (A), guanine (G), and cytosine (C), it contains the base uracil (U) instead of the thymine (T) in DNA. Since U, like T, can base-pair by hydrogen-bonding with A (Figure 6–5), the complementary base-pairing properties described for DNA in Chapters 4 and 5 apply also to RNA (in RNA, G pairs with C, and A pairs with U). It is not uncommon, however, to find other types of base pairs in RNA: for example, G pairing with U occasionally.

Despite these small chemical differences, DNA and RNA differ quite dramatically in overall structure. Whereas DNA always occurs in cells as a double-stranded helix, RNA is single-stranded. RNA chains therefore fold up into a variety of shapes, just as a polypeptide chain folds up to form the final shape of a protein (Figure 6–6). As we see later in this chapter, the ability to fold into complex three-dimensional shapes allows some RNA molecules to have structural and catalytic functions.

Transcription Produces RNA Complementary to One Strand of DNA

All of the RNA in a cell is made by DNA transcription, a process that has certain similarities to the process of DNA replication discussed in Chapter 5.

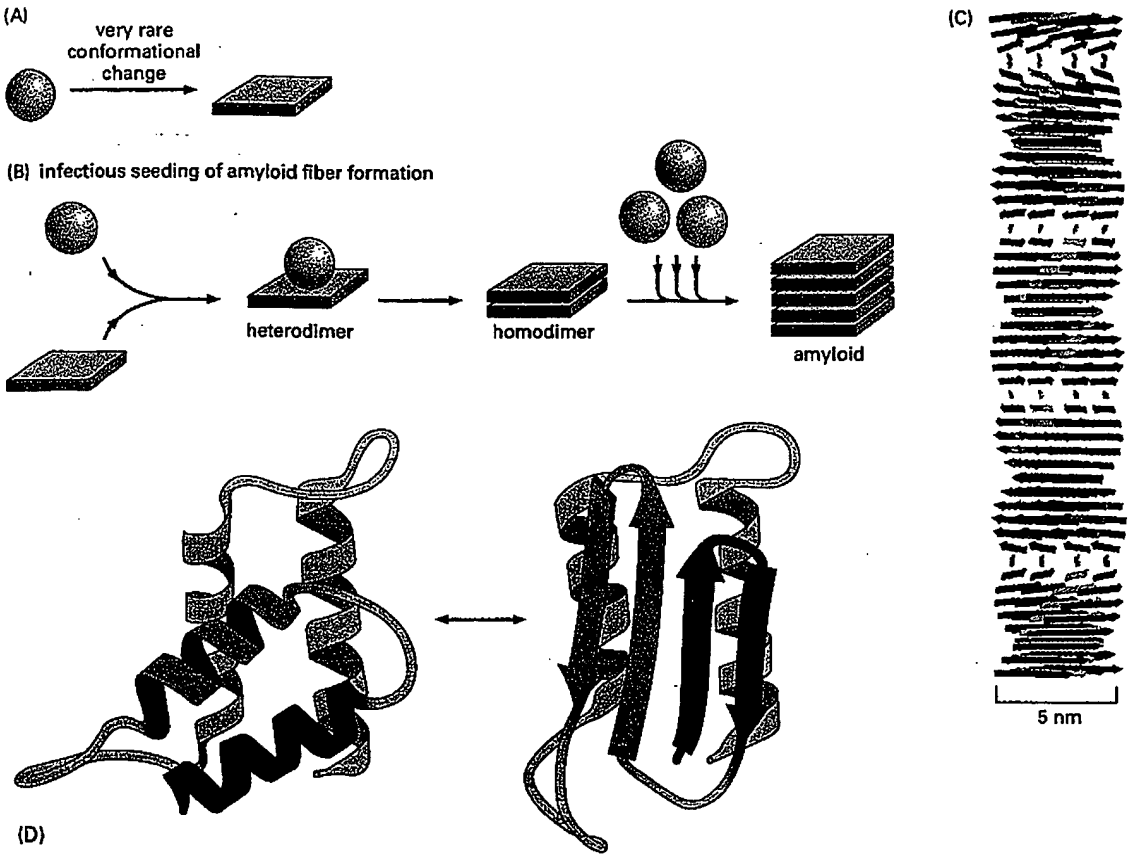


Figure 6–89 Protein aggregates that cause human disease. (A) Schematic illustration of the type of conformational change in a protein that produces material for a cross-beta filament. (B) Diagram illustrating the self-infectious nature of the protein aggregation that is central to prion diseases. PrP is highly unusual because the misfolded version of the protein, called PrP*, induces the normal PrP protein it contacts to change its conformation, as shown. Most of the human diseases caused by protein aggregation are caused by the overproduction of a variant protein that is especially prone to aggregation, but because this structure is not infectious in this way, it cannot spread from one animal to another. (C) Drawing of a cross-beta filament, a common type of protease-resistant protein aggregate found in a variety of human neurological diseases. Because the hydrogen-bond interactions in a β sheet form between polypeptide backbone atoms (see Figure 3–9), a number of different abnormally folded proteins can produce this structure. (D) One of several possible models for the conversion of PrP to PrP*, showing the likely change of two α-helices into four β-strands. Although the structure of the normal protein has been determined accurately, the structure of the infectious form is not yet known with certainty because the aggregation has prevented the use of standard structural techniques. (C, courtesy of Louise Serpell, adapted from M. Sunde et al., *J. Mol. Biol.* 273:729–739, 1997; D, adapted from S.B. Prusiner, *Trends Biochem. Sci.* 21:482–487, 1996.)

animals and humans. It can be dangerous to eat the tissues of animals that contain PrP*, as witnessed most recently by the spread of BSE (commonly referred to as the "mad cow disease") from cattle to humans in Great Britain.

Fortunately, in the absence of PrP*, PrP is extraordinarily difficult to convert to its abnormal form. Although very few proteins have the potential to misfold into an infectious conformation, a similar transformation has been discovered to be the cause of an otherwise mysterious "protein-only inheritance" observed in yeast cells.

There Are Many Steps From DNA to Protein

We have seen so far in this chapter that many different types of chemical reactions are required to produce a properly folded protein from the information contained in a gene (Figure 6–90). The final level of a properly folded protein in a cell therefore depends upon the efficiency with which each of the many steps is performed.

We discuss in Chapter 7 that cells have the ability to change the levels of their proteins according to their needs. In principle, any or all of the steps in Fig-

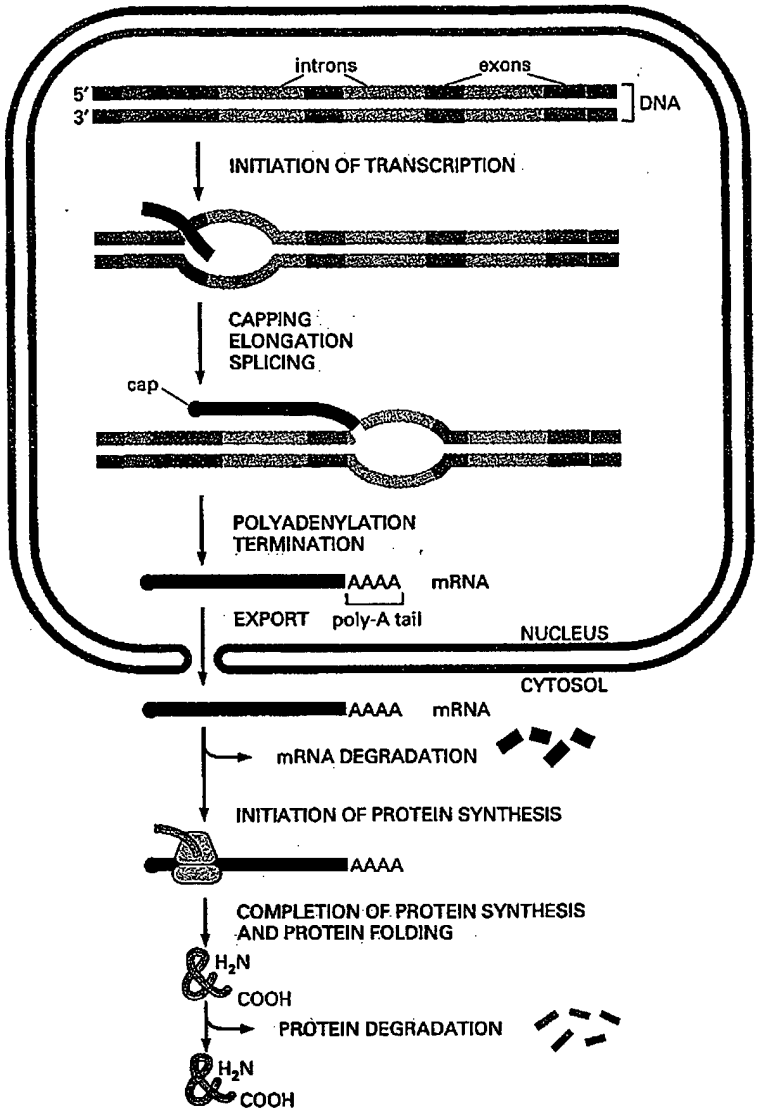


Figure 6–90 The production of a protein by a eucaryotic cell. The final level of each protein in a eucaryotic cell depends upon the efficiency of each step depicted.

ure 6–90) could be regulated by the cell for each individual protein. However, as we shall see in Chapter 7, the initiation of transcription is the most common point for a cell to regulate the expression of each of its genes. This makes sense, inasmuch as the most efficient way to keep a gene from being expressed is to block the very first step—the transcription of its DNA sequence into an RNA molecule.

Summary

The translation of the nucleotide sequence of an mRNA molecule into protein takes place in the cytoplasm on a large ribonucleoprotein assembly called a ribosome. The amino acids used for protein synthesis are first attached to a family of tRNA molecules, each of which recognizes, by complementary base-pair interactions, particular sets of three nucleotides in the mRNA (codons). The sequence of nucleotides in the mRNA is then read from one end to the other in sets of three according to the genetic code.

To initiate translation, a small ribosomal subunit binds to the mRNA molecule at a start codon (AUG) that is recognized by a unique initiator tRNA molecule. A large ribosomal subunit binds to complete the ribosome and begin the elongation phase of protein synthesis. During this phase, aminoacyl tRNAs—each bearing a specific amino acid bind sequentially to the appropriate codon in mRNA by forming complementary base pairs with the tRNA anticodon. Each amino acid is added to the C-terminal end of the growing polypeptide by means of a cycle of three sequential

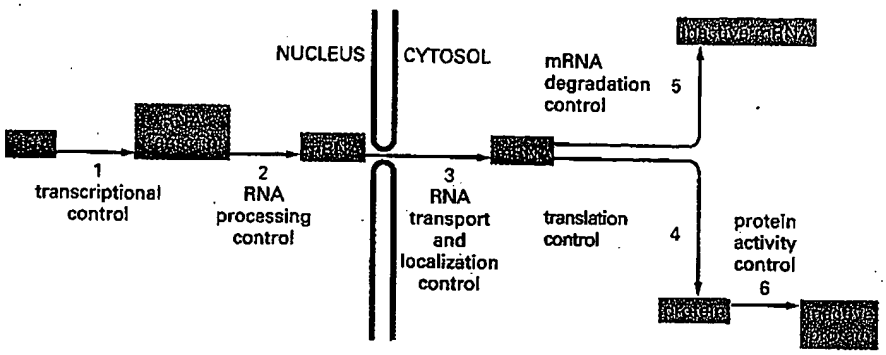


Figure 7–5 Six steps at which eucaryotic gene expression can be controlled. Controls that operate at steps 1 through 5 are discussed in this chapter. Step 6, the regulation of protein activity, includes reversible activation or inactivation by protein phosphorylation (discussed in Chapter 3) as well as irreversible inactivation by proteolytic degradation (discussed in Chapter 6).

Gene Expression Can Be Regulated at Many of the Steps in the Pathway from DNA to RNA to Protein

If differences among the various cell types of an organism depend on the particular genes that the cells express, at what level is the control of gene expression exercised? As we saw in the last chapter, there are many steps in the pathway leading from DNA to protein, and all of them can in principle be regulated. Thus a cell can control the proteins it makes by (1) controlling when and how often a given gene is transcribed (**transcriptional control**), (2) controlling how the RNA transcript is spliced or otherwise processed (**RNA processing control**), (3) selecting which completed mRNAs in the cell nucleus are exported to the cytosol and determining where in the cytosol they are localized (**RNA transport and localization control**), (4) selecting which mRNAs in the cytoplasm are translated by ribosomes (**translational control**), (5) selectively destabilizing certain mRNA molecules in the cytoplasm (**mRNA degradation control**), or (6) selectively activating, inactivating, degrading, or compartmentalizing specific protein molecules after they have been made (**protein activity control**) (Figure 7–5).

For most genes transcriptional controls are paramount. This makes sense because, of all the possible control points illustrated in Figure 7–5, only transcriptional control ensures that the cell will not synthesize superfluous intermediates. In the following sections we discuss the DNA and protein components that perform this function by regulating the initiation of gene transcription. We shall return at the end of the chapter to the additional ways of regulating gene expression.

Summary

The genome of a cell contains in its DNA sequence the information to make many thousands of different protein and RNA molecules. A cell typically expresses only a fraction of its genes, and the different types of cells in multicellular organisms arise because different sets of genes are expressed. Moreover, cells can change the pattern of genes they express in response to changes in their environment, such as signals from other cells. Although all of the steps involved in expressing a gene can in principle be regulated, for most genes the initiation of RNA transcription is the most important point of control.

DNA-BINDING MOTIFS IN GENE REGULATORY PROTEINS

How does a cell determine which of its thousands of genes to transcribe? As mentioned briefly in Chapters 4 and 6, the transcription of each gene is controlled by a regulatory region of DNA relatively near the site where transcription begins. Some regulatory regions are simple and act as switches that are thrown by a single signal. Many others are complex and act as tiny microprocessors, responding to a variety of signals that they interpret and integrate to switch the neighboring gene on or off. Whether complex or simple, these switching devices

occur in the germ line, the cell lineage that gives rise to sperm or eggs. Most of the DNA in vertebrate germ cells is inactive and highly methylated. Over long periods of evolutionary time, the methylated CG sequences in these inactive regions have presumably been lost through spontaneous deamination events that were not properly repaired. However promoters of genes that remain active in the germ cell lineages (including most housekeeping genes) are kept unmethylated, and therefore spontaneous deaminations of Cs that occur within them can be accurately repaired. Such regions are preserved in modern day vertebrate cells as CG islands. In addition, any mutation of a CG sequence in the genome that destroyed the function or regulation of a gene in the adult would be selected against, and some CG islands are simply the result of a higher than normal density of critical CG sequences.

The mammalian genome contains an estimated 20,000 CG islands. Most of the islands mark the 5′ ends of transcription units and thus, presumably, of genes. The presence of CG islands often provides a convenient way of identifying genes in the DNA sequences of vertebrate genomes.

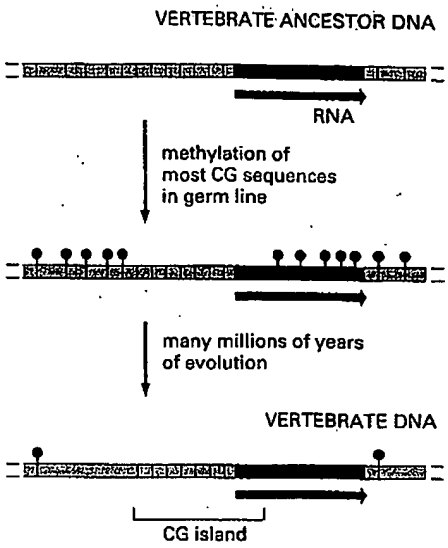


Figure 7–86 A mechanism to explain both the marked overall deficiency of CG sequences and their clustering into CG islands in vertebrate genomes. A *black line* marks the location of a CG dinucleotide in the DNA sequence, while a *red* "lollipop" indicates the presence of a methyl group on the CG dinucleotide. CG sequences that lie in regulatory sequences of genes that are transcribed in germ cells are unmethylated and therefore tend to be retained in evolution. Methylated CG sequences, on the other hand, tend to be lost through deamination of 5-methyl C to T, unless the CG sequence is critical for survival.

Summary

The many types of cells in animals and plants are created largely through mechanisms that cause different genes to be transcribed in different cells. Since many specialized animal cells can maintain their unique character through many cell division cycles and even when grown in culture, the gene regulatory mechanisms involved in creating them must be stable once established and heritable when the cell divides. These features endow the cell with a memory of its developmental history. Bacteria and yeasts provide unusually accessible model systems in which to study gene regulatory mechanisms. One such mechanism involves a competitive interaction between two gene regulatory proteins, each of which inhibits the synthesis of the other; this can create a flip-flop switch that switches a cell between two alternative patterns of gene expression. Direct or indirect positive feedback loops, which enable gene regulatory proteins to perpetuate their own synthesis, provide a general mechanism for cell memory. Negative feedback loops with programmed delays form the basis for cellular clocks.

In eucaryotes the transcription of a gene is generally controlled by combinations of gene regulatory proteins. It is thought that each type of cell in a higher eucaryotic organism contains a specific combination of gene regulatory proteins that ensures the expression of only those genes appropriate to that type of cell. A given gene regulatory protein may be active in a variety of circumstances and typically is involved in the regulation of many genes.

In addition to diffusible gene regulatory proteins, inherited states of chromatin condensation are also used by eucaryotic cells to regulate gene expression. An especially dramatic case is the inactivation of an entire X chromosome in female mammals. In vertebrates DNA methylation also functions in gene regulation, being used mainly as a device to reinforce decisions about gene expression that are made initially by other mechanisms. DNA methylation also underlies the phenomenon of genomic imprinting in mammals, in which the expression of a gene depends on whether it was inherited from the mother or the father.

POSTTRANSCRIPTIONAL CONTROLS

In principle, every step required for the process of gene expression could be controlled. Indeed, one can find examples of each type of regulation, although any one gene is likely to use only a few of them. Controls on the initiation of gene transcription are the predominant form of regulation for most genes. But other controls can act later in the pathway from DNA to protein to modulate the amount of gene product that is made. Although these **posttranscriptional controls**, which operate after RNA polymerase has bound to the gene's promoter and begun RNA synthesis, are less common than *transcriptional control*, for many genes they are crucial.

CHAPTER 29

Regulation of transcription

Genes VII (1997) CH 29, pp. 847-848.
Benjamin Lewin

The phenotypic differences that distinguish the various kinds of cells in a higher eukaryote are largely due to differences in the expression of genes that code for proteins, that is, those transcribed by RNA polymerase II. In principle, the expression of these genes might be regulated at any one of several stages. The concept of the "level of control" implies that gene expression is not necessarily an automatic process once it has begun. It could be regulated in a gene-specific way at any one of several sequential steps. We can distinguish (at least) five potential control points, forming the series:

Activation of gene structure
↓
Initiation of transcription
↓
Processing the transcript
↓
Transport to cytoplasm
↓
Translation of mRNA

The existence of the first step is implied by the discovery that genes may exist in either of two structural conditions. Relative to the state of most of the genome, genes are found in an "active" state in the cells in which they are expressed (see Chapter 27). The change of structure is distinct from the act of transcription, and indicates that the gene is "transcribable." This suggests that acquisition of the "active" structure must be the first step in gene expression.

Transcription of a gene in the active state is controlled at the stage of initiation, that is, by the interaction of RNA polymerase with its promoter. This is now becoming susceptible to analysis in the *in vitro* systems (see Chapter 28). For most genes, this is a major control point; probably it is the most common level of regulation.

There is at present no evidence for control at subsequent stages of transcription in eukaryotic cells, for example, via antitermination mechanisms.

The primary transcript is modified by capping at the 5' end, and usually also by polyadenylation at the 3' end. Introns must be spliced out from the transcripts of interrupted genes. The mature RNA must be exported from the nucleus to the cytoplasm. Regulation of gene expression by selection of sequences at the level of nuclear RNA might involve any or all of these stages, but the one for which we have most evidence concerns changes in splicing: some genes are expressed by means of alternative splicing patterns whose regulation controls the type of protein product (see Chapter 30).

Finally, the translation of an mRNA in the cytoplasm can be specifically controlled. There is little evidence for the employment of this mechanism in adult somatic cells, but it does occur in some embryonic situations, as described in Chapter 7. The mechanism is presumed to involve the blocking of initiation of translation of some mRNAs by specific protein factors.

But having acknowledged that control of gene expression can occur at multiple stages, and that production of RNA cannot inevitably be equated with production of protein, it is clear

that the overwhelming majority of regulatory events occur at the initiation of transcription. Regulation of tissue-specific gene transcription lies at the heart of eukaryotic differentiation; indeed, we see examples in Chapter 38 in which proteins that regulate embryonic development prove to be transcription factors. A regulatory transcription factor serves to provide common control of a large number of target genes, and we seek to answer two questions about this mode of regulation: what identifies the common target genes to the transcription factor; and how is the activity of the transcription factor itself regulated in response to intrinsic or extrinsic signals?

Response elements identify genes under common regulation

The principle that emerges from characterizing groups of genes under common control is that *they share a promoter element that is recognized by a regulatory transcription factor.* An element that causes a gene to respond to such a factor is called a response element; examples are the **HSE** (heat shock response element), **GRE** (glucocorticoid response element), **SRE** (serum response element).

The properties of some inducible transcription factors and the elements that they recognize are summarized in Table 29.1. Response elements have the same general characteristics as upstream elements of promoters or enhancers. They contain short consensus sequences, and copies of the response elements found in different genes are closely related, but not necessarily identical. The region bound by the factor extends for a short distance on either side of the consensus sequence. In promoters, the elements are not present at fixed distances from the startpoint, but are usually <200 bp upstream of it. The presence of a single element usually is sufficient to confer the regulatory response, but sometimes there are multiple copies.

Table 29.1 Inducible transcription factors bind to response elements that identify groups of promoters or enhancers subject to coordinate control.

Regulatory Agent	Module	Consensus	Factor
Heat shock	HSE	CNNGAANNTCCNNG	HSTF
Glucocorticoid	GRE	TGGTACAAATGTTCT	Receptor
Phorbol ester	TRE	TGACTCA	AP1
Serum	SRE	CCATATTAGG	SRF

Response elements may be located in promoters or in enhancers. Some types of elements are typically found in one rather than the other: usually an HSE is found in a promoter, while a GRE is found in an enhancer. We assume that all response elements function by the same general principle. *A gene is regulated by a sequence at the promoter or enhancer that is recognized by a specific protein. The protein functions as a transcription factor needed for RNA polymerase to initiate. Active protein is available only under conditions when the gene is to be expressed; its absence means that the promoter is not activated by this particular circuit.*

An example of a situation in which many genes are controlled by a single factor is provided by the heat shock response. This is common to a wide range of prokaryotes and eukaryotes and involves multiple controls of gene expression: an increase in temperature turns off transcription of some genes, turns on transcription of the heat shock genes, and causes changes in the translation of mRNAs. The control of the heat shock genes illustrates the differences between prokaryotic and eukaryotic modes of control. In bacteria, a new sigma factor is synthesized that directs RNA polymerase holoenzyme to recognize an alter-

World Journal of Surgical Oncology

BioMed Central

Research

Open Access

Prostate stem cell antigen (PSCA) expression in human prostate cancer tissues and its potential role in prostate carcinogenesis and progression of prostate cancer

Zhao Zhigang*[1] and Shen Wenlv[2]

Address: [1]Department of Urology, Shantou University Medical College, Shantou, Guangdong, China and [2]Department of Urology, No 2. Affiliated Hospital of Shantou University Medical College, Shantou, Guangdong, China

Email: Zhao Zhigang* - zgzhaodr@yahoo.com.cn; Shen Wenlv - wlshen99@hotmail.com

* Corresponding author

Published: 10 May 2004

World Journal of Surgical Oncology 2004, 2:13

Received: 30 March 2004

Accepted: 10 May 2004

This article is available from: http://www.wjso.com/content/2/1/13

Abstract

Background: Prostate stem cell antigen (PSCA) is a recently defined homologue of the Thy-1/Ly-6 family of glycosylphosphatidylinositol (GPI)-anchored cell surface antigens. The purpose of the present study was to examine the expression status of PSCA protein and mRNA in clinical specimens of human prostate cancer (Pca) and to validate it as a potential molecular target for diagnosis and treatment of Pca.

Materials and Methods: Immunohistochemical (IHC) and *in situ* hybridization (ISH) analyses of PSCA expression were simultaneously performed on paraffin-embedded sections from 20 benign prostatic hyperplasia (BPH), 20 prostatic intraepithelial neoplasm (PIN) and 48 prostate cancer (Pca) tissues, including 9 androgen-independent prostate cancers. The level of PSCA expression was semiquantitatively scored by assessing both the percentage and intensity of PSCA-positive staining cells in the specimens. Then compared PSCA expression between BPH, PIN and Pca tissues and analysed the correlations of PSCA expression level with pathological grade, clinical stage and progression to androgen-independence in Pca.

Results: In BPH and low grade PIN, PSCA protein and mRNA staining were weak or negative and less intense and uniform than that seen in HGPIN and Pca. There were moderate to strong PSCA protein and mRNA expression in 8 of 11 (72.7%) HGPIN and in 40 of 48 (83.4%) Pca specimens examined by IHC and ISH analyses, with statistical significance compared with BPH (20%) and low grade PIN (22.2%) samples ($p < 0.05$, respectively). The expression level of PSCA increased with high Gleason grade, advanced stage and progression to androgen-independence ($p < 0.05$, respectively). In addition, IHC and ISH staining showed a high degree of correlation between PSCA protein and mRNA overexpression.

Conclusions: Our data demonstrate that PSCA as a new cell surface marker is overexpressed by a majority of human Pca. PSCA expression correlates positively with adverse tumor characteristics, such as increasing pathological grade (poor cell differentiation), worsening clinical stage and androgen-independence, and speculatively with prostate carcinogenesis. PSCA protein overexpression results from upregulated transcription of PSCA mRNA. PSCA may have prognostic utility and may be a promising molecular target for diagnosis and treatment of Pca.

Introduction

Prostate cancer (Pca) is the second leading cause of cancer-related death in American men and is becoming a common cancer increasing in China. Despite recently great progress in the diagnosis and management of localized disease, there continues to be a need for new diagnostic markers that can accurately discriminate between indolent and aggressive variants of Pca. There also continues to be a need for the identification and characterization of potential new therapeutic targets on Pca cells. Current diagnostic and therapeutic modalities for recurrent and metastatic Pca have been limited by a lack of specific target antigens of Pca.

Although a number of prostate-specific genes have been identified (i.e. prostate specific antigen, prostatic acid phosphatase, glandular kallikrein 2), the majority of these are secreted proteins not ideally suited for many immunological strategies. So, the identification of new cell surface antigens is critical to the development of new diagnostic and therapeutic approaches to the management of Pca.

Reiter RE et al [1] reported the identification of prostate stem cell antigen (PSCA), a cell surface antigen that is predominantly prostate specific. The PSCA gene encodes a 123 amino acid glycoprotein, with 30% homology to stem cell antigen 2 (Sca 2). Like Sca-2, PSCA also belongs to a member of the Thy-1/Ly-6 family and is anchored by a glycosylphosphatidylinositol (GPI) linkage. mRNA *in situ* hybridization (ISH) localized PSCA expression in normal prostate to the basal cell epithelium, the putative stem cell compartment of prostatic epithelium, suggesting that PSCA may be a marker of prostate stem/progenitor cells.

In order to examine the status of PSCA protein and mRNA expression in human Pca and validate it as a potential diagnostic and therapeutic target for Pca, we used immunohistochemistry (IHC) and *in situ* hybridization (ISH) simultaneously, and conducted PSCA protein and mRNA expression analyses in paraffin-embedded tissue specimens of benign prostatic hyperplasia (BPH, n = 20), prostate intraepithelial neoplasm (PIN, n = 20) and prostate cancer (Pca, n = 48). Furthermore, we evaluated the possible correlation of PSCA expression level with Pca tumorigenesis, grade, stage and progression to androgen-independence.

Materials and methods

Tissue samples

All of the clinical tissue specimens studied herein were obtained from 80 patients of 57–84 years old by prostatectomy, transurethral resection of prostate (TURP) or biopsies. The patients were classified as 20 cases of BPH, 20 cases of PIN, 40 cases of primary Pca, including 9 patients with recurrent Pca and a history of androgen ablation therapy (orchiectomy and/or hormonal therapy), who were referred to as androgen-independent prostate cancers. Eight specimens were harvested from these androgen-independent Pca patients prior to androgen ablation treatment. Each tissue sample was cut into two parts, one was fixed in 10% formalin for IHC and the other treated with 4% paraformaldehyde/0.1 M PBS PH 7.4 in 0.1% DEPC for 1 h for ISH analysis, and then embedded in paraffin. All paraffin blocks examined were then cut into 5 μm sections and mounted on the glass slides specific for IHC and ISH respectively in the usual fashion. H&E-stained section of each Pca was evaluated and assigned a Gleason score by the experienced urological pathologist at our institution based on the criteria of Gleason score [2]. The Gleason sums are summarized in Table 1. Clinical staging was performed according to Jewett-whitmore-prout staging system, as shown in Table 2. In the category of PIN, we graded the specimens into two groups, i.e. low grade PIN (grade I - II) and high grade PIN (HGPIN, grade III) on the basis of literatures [3,4].

Immunohistochemical (IHC) analysis

Briefly, tissue sections were deparaffinized, dehydrated, and subjected to microwaving in 10 mmol/L citrate buffer, PH 6.0 (Boshide, Wuhan, China) in a 900 W oven for 5 min to induce epitope retrieval. Slides were allowed to cool at room temperature for 30 min. A primary mouse antibody specific to human PSCA (Boshide, Wuhan, China) with a 1:100 dilution was applied to incubate with the slides at room temperature for 2 h. Labeling was detected by sequentially adding biotinylated secondary antibodies and strepavidin-peroxidase, and localized using 3,3'-diaminobenzidine reaction. Sections were then counterstained with hematoxylin. Substitution of the primary antibody with phosphate-buffered-saline (PBS) served as a negative-staining control.

mRNA in situ hybridization (ISH)

Five-μm-thick tissue sections were deparaffinized and dehydrated, then digested in pepsin solution (4 mg/ml in 3% citric acid) for 20 min at 37.5°C, and further processed for ISH. Digoxigenin-labeled sense and antisense human PSCA RNA probes (obtained from Boshide, Wuhan, China) were hybridized to the sections at 48°C overnight. The posthybridization wash with a high stringency was performed sequentially at 37°C in 2 × standard saline citrate (SSC) for 10 min, in 0.5 × SSC for 15 min and in 0.2 × SSC for 30 min. The slides were then incubated to biotinylated mouse anti-digoxigenin antibody at 37.5°C for 1 h followed by washing in 1 × PBS for 20 min at room temperature, and then to strepavidin-peroxidase at 37.5°C for 20 min followed by washing in 1 × PBS for 15 min at room temperature. Subsequently, the slides were developed with diaminobenzidine and then coun-

Table 1: Correlation of PSCA expression with Gleason score

	Intensity × frequency	
Gleason score	0–6 (%)	9 (%)
2–4	5 (83)	1 (17)
5–7	19 (79)	5 (21)
8–10	5 (28)	13 (72)

Table 2: Correlation of PSCA expression with clinical stage

	Intensity × frequency	
Tumor stage	0–6 (%)	9 (%)
≤B	27 (67.5)	13 (32.5)
≥C	2 (25)	6 (75)

terstained with hematoxylin to localize the hybridization signals. Sections hybridized with the sense control probes routinely did not show any specific hybridization signal above background. All slides were hybridized with PBS to substitute for the probes as a negative control.

Scoring methods

To determine the correlation between the results of PSCA immunostaining and mRNA *in situ* hybridization, the same scoring manners are taken in the present study for PSCA protein staining by IHC and PSCA mRNA staining by ISH. Each slide was read and scored by two independently experienced urological pathologists using Olympus BX-41 light microscopes. The evaluation was done in a blinded fashion. For each section, five areas of similar grade were analyzed semiquantitatively for the fraction of cells staining. Fifty percent of specimens were randomly chosen and rescored to determine the degree of interobserver and intraobserver concordance. There was greater than 95% intra- and interobserver agreement.

The intensity of PSCA expression evaluated microscopically was graded on a scale of 0 to 3+ with 3 being the highest expression observed (0, no staining; 1+, mildly intense; 2+, moderately intense; 3+, severely intense). The staining density was quantified as the percentage of cells staining positive for PSCA with the primary antibody or hybridization probe, as follows: 0 = no staining; 1 = positive staining in <25% of the sample; 2 = positive staining in 25%–50% of the sample; 3 = positive staining in >50% of the sample. Intensity score (0 to 3+) was multiplied by the density score (0–3) to give an overall score of 0–9 [1,5]. In this way, we were able to differentiate specimens that may have had focal areas of increased staining from those that had diffuse areas of increased staining [6]. The overall score for each specimen was then categorically assigned to one of the following groups: 0 score, negative expression; 1–2 scores, weak expression; 3–6 scores, moderate expression; 9 score, strong expression.

Statistical analysis

Intensity and density of PSCA protein and mRNA expression in BPH, PIN and Pca tissues were compared using the Chi-square and Student's *t*-test. Univariate associations between PSCA expression and Gleason score, clinical stage and progression to androgen-independence were calculated using Fisher's Exact Test. For all analyses, $p < 0.05$ was considered statistically significant.

Results

PSCA expression in BPH

In general, PSCA protein and mRNA were expressed weakly in individual samples of BPH. Some areas of prostate expressed weak levels (composite score 1–2), whereas other areas were completely negative (composite score 0). Four cases (20%) of BPH had moderate expression of PSCA protein and mRNA (composite score 4–6) by IHC and ISH. In 2/20 (10%) BPH specimens, PSCA mRNA expression was moderate (composite score 3–6), but PSCA protein expression was weak (composite score

2) in one and negative (composite score 0) in the other. PSCA expression was localized to the basal and secretory epithelial cells, and prostatic stroma was almost negative staining for PSCA protein and mRNA in all cases examined.

PSCA expression in PIN

In this study, we detected weak or negative expression of PSCA protein and mRNA (≤2 scores) in 7 of 9 (77.8%) low grade PIN and in 2 of 11 (18.2%) HGPIN, and moderate expression (3–6 scores) in the rest 2 low grade PIN and 5 of 11 (45.5%) HGPIN. One HGPIN with moderate PSCA mRNA expression (6 score) was found weak staining for PSCA protein (2 score) by IHC. Strong PSCA protein and mRNA expression (9 score) were detected in the remaining 3 of 11 (27.3%) HGPIN. There was a statistically significant difference of PSCA protein and mRNA expression levels observed between HGPIN and BPH ($p < 0.05$), but no statistical difference reached between low grade PIN and BPH ($p > 0.05$).

PSCA expression in Pca

In order to determine if PSCA protein and mRNA can be detected in prostate cancers and if PSCA expression levels are increased in malignant compared with benign glands, Forty-eight paraffin-embedded Pca specimens were analysed by IHC and ISH. It was shown that 19 of 48 (39.6%) Pca samples stained very strongly for PSCA protein and mRNA with a score of 9 and another 21 (43.8%) specimens displayed moderate staining with scores of 4–6 (Figure 1). In addition, 4 specimens with moderate to strong PSCA mRNA expression (scores of 4–9) had weak protein staining (a score of 2) by IHC analyses. Overall, Pca expressed a significantly higher level of PSCA protein and mRNA than any other specimen category in this study ($p < 0.05$, compared with BPH and PIN respectively). The result demonstrates that PSCA protein and mRNA are overexpressed by a majority of human Pca.

Correlation of PSCA expression with Gleason score in Pca

Using the semi-quantitative scoring method as described in Materials and Methods, we compared the expression level of PSCA protein and mRNA with Gleason grade of Pca, as shown in Table 1. Prostate adenocarcinomas were graded by Gleason score as 2–4 scores = well-differentiation, 5–7 scores = moderate-differentiation and 8–10 scores = poor-differentiation [7]. Seventy-two percent of Gleason scores 8–10 prostate cancers had very strong staining of PSCA compared to 21% with Gleason scores 5–7 and 17% with 2–4 respectively, demonstrating that poorly differentiated Pca had significantly stronger expression of PSCA protein and mRNA than moderately and well differentiated tumors ($p < 0.05$). As depicted in Figure 1, IHC and ISH analyses showed that PSCA protein and mRNA expression in several cases of poorly differentiated Pca were particularly prominent, with more intense and uniform staining. The results indicate that PSCA expression increases significantly with higher tumor grade in human Pca.

Correlation of PSCA expression with clinical stage in Pca

With regards to PSCA expression in every stage of Pca, we showed the results in Table 2. Seventy-five percent of locally advanced and node positive cancers (i.e. C-D stages) expressed statistically high levels of PSCA versus 32.5% that were organ confined (i.e. A-B stages) ($p < 0.05$). The data demonstrate that PSCA expression increases significantly with advanced tumor stage in human Pca.

Correlation of PSCA expression with androgen-independent progression of Pca

All 9 specimens of androgen-independent prostate cancers stained positive for PSCA protein and mRNA. Eight specimens were obtained from patients managed prior to androgen ablation therapy. Seven of eight (87.5%) of these androgen-independent prostate cancers were in the strongest staining category (score = 9), compared with three out of eight (37.5%) of patients with androgen-dependent cancers ($p < 0.05$). The results demonstrate that PSCA expression increases significantly with progression to androgen-independence of human Pca.

It is evident from the results above that within a majority of human prostate cancers the level of PSCA protein and mRNA expression correlates significantly with increasing grade, worsening stage and progression to androgen-independence.

Correlation of PSCA immunostaining and mRNA in situ hybridization

In all 88 specimens surveyed herein, we compared the results of PSCA IHC staining with mRNA ISH analysis. Positive staining areas and its intensity and density scores evaluated by IHC were identical to those seen by ISH in 79 of 88 (89.8%) specimens (18/20 BPH, 19/20 PIN and 42/48 Pca respectively). Importantly, 27/27 samples with PSCA mRNA composite scores of 0–2, 32/36 samples with scores of 3–6 and 22/24 samples with a score of 9 also had PSCA protein expression scores of 0–2, 3–6 and 9 respectively. However, in 5 samples with PSCA mRNA overall scores of 3–6 and in 2 with scores of 9 there were less or negative PSCA protein expression (i.e. scores of 0–4), suggesting that this may reflect posttranscriptional modification of PSCA or that the epitopes recognized by PSCA mAb may be obscured in some cancers. The data demonstrate that the results of PSCA immunostaining were consistent with those of mRNA ISH analysis, showing a high degree of correlation between PSCA protein and mRNA expression.

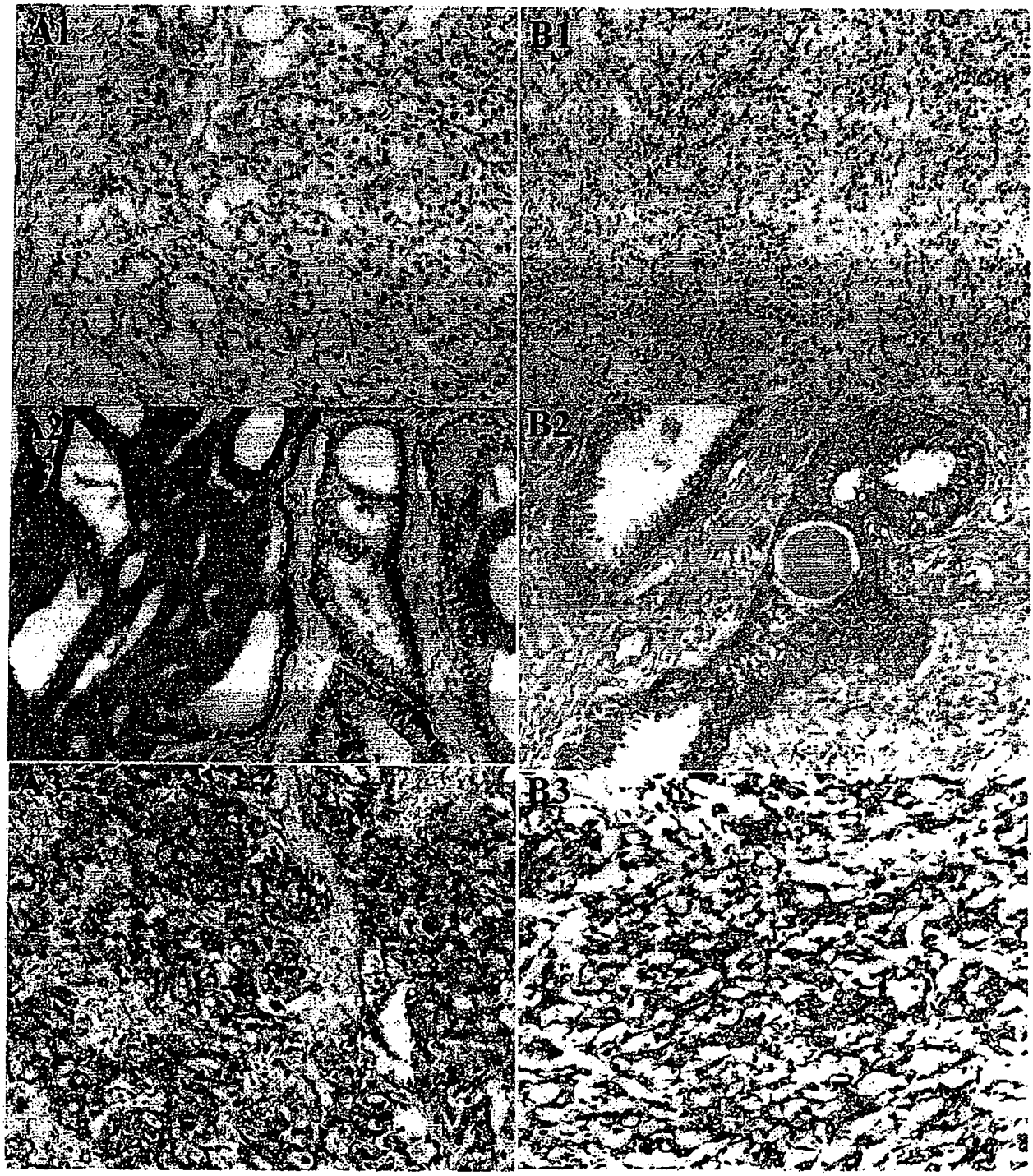


Figure 1

Representatives of PSCA IHC and ISH staining in Pca (A. IHC staining, B. ISH staining, ×200 magnification). A_1, B_1: negative control of IHC and ISH. PBS replacing the primary antibody (A_1) and hybridization with a sense PSCA probe (B_1) showed no background staining. A_2, B_2: a moderately differentiated Pca (Gleason score = 3+3 = 6) with moderate staining (composite score = 6) in all malignant cells; A_2: IHC shows not only cell surface but also apparent cytoplasmic staining of PSCA protein. A_3, B_3: a poorly differentiated Pca (Gleason score = 4+4 = 8) with very strong staining (composite score = 9) in all malignant cells.

Discussion

PSCA is homologous to a group of cell surface proteins that mark the earliest phase of hematopoietic development. PSCA mRNA expression is prostate-specific in normal male tissues and is highly up-regulated in both androgen-dependent and-independent Pca xenografts (LAPC-4 tumors). We hypothesize that PSCA may play a role in Pca tumorigenesis and progression, and may serve as a target for Pca diagnosis and treatment. In this study, IHC and ISH showed that in general there were weak or absent PSCA protein and mRNA expression in BPH and low grade PIN tissues. However, PSCA protein and mRNA are widely expressed in HGPIN, the putative precursor of invasive Pca, suggesting that up-regulation of PSCA is an early event in prostate carcinogenesis. Recently, Reiter RE et al [1], using ISH analysis, reported that 97 of 118 (82%) HGPIN specimens stained strongly positive for PSCA mRNA. A very similar finding was seen on mouse PSCA (mPSCA) expression in mouse HGPIN tissues by Tran C. P et al [8]. These data suggest that PSCA may be a new marker associated with transformation of prostate cells and tumorigenesis.

We observed that PSCA protein and mRNA are highly expressed in a large percentage of human prostate cancers, including advanced, poorly differentiated, androgen-independent and metastatic cases. Fluorescence-activated cell sorting and confocal/ immunofluorescent studies demonstrated cell surface expression of PSCA protein in Pca cells [9]. Our IHC expression analysis of PSCA shows not only cell surface but also apparent cytoplasmic staining of PSCA protein in Pca specimens (Figure 1). One possible explanation for this is that anti-PSCA antibody can recognize PSCA peptide precursors that reside in the cytoplasm. Also, it is possible that the positive staining that appears in the cytoplasm is actually from the overlying cell membrane [5]. These data seem to indicate that PSCA is a novel cell surface marker for human Pca.

Our results show that elevated level of PSCA expression correlates with high grade (i.e. poor differentiation), increased tumor stage and progression to androgen-independence of Pca. These findings support the original IHC analyses by Gu Z et al [9], who reported that PSCA protein expressed in 94% of primary Pca and the intensity of PSCA protein expression increased with tumor grade, stage and progression to androgen-independence. Our results also collaborate the recent work of Han KR et al [10], in which the significant association between high PSCA expression and adverse prognostic features such as high Gleason score, seminal vesicle invasion and capsular involvement in Pca was found. It is suggested that PSCA overexpression may be an adverse predictor for recurrence, clinical progression or survival of Pca. Hara H et al [11] used RT-PCR detection of PSA, PSMA and PSCA in 1 ml of peripheral blood to evaluate Pca patients with poor prognosis. The results showed that among 58 PCa patients, each PCR indicated the prognostic value in the hierarchy of PSCA>PSA>PSMA RT-PCR, and extraprostatic cases with positive PSCA PCR indicated lower disease-progression-free survival than those with negative PSCA PCR, demonstrating that PSCA can be used as a prognostic factor. Dubey P et al [12] reported that elevated numbers of PSCA + cells correlate positively with the onset and development of prostate carcinoma over a long time span in the prostates of the TRAMP and PTEN +/- models compared with its normal prostates. Taken together with our present findings, in which PSCA is overexpressed from HGPIN to almost frank carcinoma, it is reasonable and possible to use increased PSCA expression level or increased numbers of PSCA-positive cells in the prostate samples as a prognostic marker to predict the potential onset of this cancer. These data raise the possibility that PSCA may have diagnostic utility or clinical prognostic value in human Pca.

The cause of PSCA overexpression in Pca is not known. One possible mechanism is that it may result from PSCA gene amplification. In humans, PSCA is located on chromosome 8q24.2 [1], which is often amplified in metastatic and recurrent Pca and considered to indicate a poor prognosis [13-15]. Interestingly, PSCA is in close proximity to the c-myc oncogene, which is amplified in >20% of recurrent and metastatic prostate cancers [16,17]. Reiter RE et al [18] reported that PSCA and MYC gene copy numbers were co-amplified in 25% of tumors (five out of twenty), demonstrating that PSCA overexpression is associated with PSCA and MYC coamplification in Pca. Gu Z et al [9] recently reporteted that in 102 specimens available to compare the results of PSCA immunostaining with their previous mRNA ISH analysis, 92 (90.2%) had identically positive areas of PSCA protein and mRNA expression. Taken together with our findings, in which we detected moderate to strong expression of PSCA protein and mRNA in 34 of 40 (85%) Pca specimens examined simultaneously by IHC and ISH analyses, it is demonstrated that PSCA protein and mRNA overexpressed in human Pca, and that the increased protein level of PSCA was resulted from the upregulated transcription of its mRNA.

At present, the regulation mechanisms of human PSCA expression and its biological function are yet to be elucidated. PSCA expression may be regulated by multiple factors [18]. Watabe T et al [19] reported that transcriptional control is a major component regulating PSCA expression levels. In addition, induction of PSCA expression may be regulated or mediated through cell-cell contact and protein kinase C (PKC) [20]. Homologues of PSCA have diverse activities, and have themselves been involved in

carcinogenesis. Signalling through SCA-2 has been demonstrated to prevent apoptosis in immature thymocytes [21]. Thy-1 is involved in T cell activation and transducts signals through src-like tyrosine kinases [22]. Ly-6 genes have been implicated both in tumorigenesis and in cell-cell adhesion [23-25]. Cell-cell or cell-matrix interaction is critical for local tumor growth and spread to distal sites. From its restricted expression in basal cells of normal prostate and its homology to SCA-2, PSCA may play a role in stem/progenitor cell function, such as self-renewal (i.e. anti-apoptosis) and/or proliferation [1]. Taken together with the results in the present study, we speculate that PSCA may play a role in tumorigenesis and clinical progression of Pca through affecting cell transformation and proliferation. From our results, it is also suggested that PSCA as a new cell surface antigen may have a number of potential uses in the diagnosis, therapy and clinical prognosis of human Pca. PSCA overexpression in prostate biopsies could be used to identify patients at high risk to develop recurrent or metastatic disease, and to discriminate cancers from normal glands in prostatectomy samples. Similarly, the detection of PSCA-overexpressing cells in bone marrow or peripheral blood may identify and predict metastatic progression better than current assays, which identify only PSA-positive or PSMA-positive prostate cells.

In summary, we have shown in this study that PSCA protein and mRNA are maintained in expression from HGPIN through all stages of Pca in a majority of cases, which may be associated with prostate carcinogenesis and correlate positively with high tumor grade (poor cell differentiation), advanced stage and androgen-independent progression. PSCA protein overexpression is due to the upregulation of its mRNA transcription. The results suggest that PSCA may be a promising molecular marker for the clinical prognosis of human Pca and a valuable target for diagnosis and therapy of this tumor.

Competing interests

None declared.

References

1. Reiter RE, Gu Z, Watabe T, Thomas G, Szigeti K, David E, Wahl M, Nisitani S, Yamashiro J, Le Beau MM, Loda M, Witte ON: **Prostate stem cell antigen: a cell surface marker overexpressed in prostate cancer.** *Proc Natl Acad Sci USA* 1998, **95**:1735-1740.
2. Gleason DF: **Histologic grading and clinical staging of prostatic carcinoma.** In: *Urologic Pathology: The Prostate* Edited by: Tannebaum M. Philadelphia, Lea & Febiger; 1977:171-197.
3. Brawer MK: **Prostatic intraepithelial neoplasia: a premalignant lesion.** *Hum Pathol* 1992, **23**:242-248.
4. Amin MB, Ro JY, Ayala AC: **Prostatic intraepithelial neoplasia: relationship to adenocarcinoma of prostate.** *Pathol Annu* 1994, **29**:1-30.
5. Amara N, Palapattu GS, Schrage M, Gu Z, Thomas GV, Dorey F, Said J, Reiter RE: **Prostate stem cell antigen is overexpressed in human transitional cell carcinoma.** *Cancer Res* 2001, **61**:4660-4665.
6. Hanas JS, Lerner MR, Lightfoot SA, Raczkowski C, Kastens DJ, Brackett DJ, Postier RG: **Expression of the cyclin-dependent kinase inhibitor p21 (WAF1/CIP1) and p53 tumor suppressor in dysplastic progression and adenocarcinoma in Barrett esophagus.** *Cancer (Phila)* 1999, **86**:756-763.
7. Egevad L, Gramfors T, Karlberg L: **Prognostic value of the Gleason score in prostate cancer.** *BJU Int* 2002, **89**:538-542.
8. Tran CP, Lin C, Yamashiro J, Reiter RE: **Prostate stem cell antigen is a marker of late intermediate prostate epithelial cells.** *Mol Cancer Res* 2002, **1**:113-121.
9. Gu Z, Thomas G, Yamashiro J, Shintaku IP, Dorey F, Raitano A, Witte ON, Said JW, Loda M, Reiter RE: **Prostate stem cell antigen (PSCA) expression increases with high Gleason score, advanced stage and bone metastasis in prostate cancer.** *Oncogene* 2000, **19**:1288-1296.
10. Han KR, Seligson DB, Liu X, Horvath S, Shintaku PI, Thomas GV, Said JW, Reiter RE: **Prostate stem cell antigen expression is associated with gleason score, seminal vesicle invasion and capsular invasion in prostate cancer.** *J Urol* 2004, **171**:1117-1121.
11. Hara H, Kasahara T, Kawasaki T, Bilim V, Obara K, Takahashi K, Tomita Y: **Reverse Transcription-Polymerase Chain Reaction Detection of Prostate-specific Antigen, Prostate-specific Membrane Antigen, and Prostate Stem Cell Antigen in One Milliliter of Peripheral Blood.** *Clin Cancer Res* 2002, **8**:1794-1799.
12. Dubey P, Wu H, Reiter RE, Witte ON: **Alternative pathways to prostate carcinoma activate prostate stem cell antigen expression.** *Cancer Res* 2001, **61**:3256-3261.
13. Visa korpi T, Kallioniemi AH, Syvanen AC, Hyytinen ER, Karhu R, Tammela T, Isola JJ, Kallioniemi OP: **Genetic changes in primary and recrrent prostate cancer by comparative genomic hybridization.** *Cancer Res* 1995, **55**:342-347.
14. Sato K, Qian J, Slezak JM, Lieber MM, Bostwick DG, Bergstralh EJ, Jenkins RB: **Clinical significance of alterations of chromosome 8 in high-grade, advanced, nonmetastatic prostate carcinoma.** *J Natl Cancer Inst* 1999, **91**:1574-1580.
15. Van Den Berg C, Guan XY, Von Hoff D, Jenkins R, Bittner , Griffin C, Kallioniemi O, Visakorpi , McGill , Herath J, Epstein J, Sarosdy M, Meltzer P, Trent J: **DNA sequence amplification in human prostate cancer identified by chromosome microdissection: potential prognostic implications.** *Clin Cancer Res* 1995, **1**:11-18.
16. Jenkins RB, Qian J, Lieber MM, Bostwick DG: **Detection of c-myc oncogene amplification and chromosomal anomalies in metastatic prostatic carcinoma by fluorescence in situ hybridization.** *Cancer Res* 1997, **57**:524-531.
17. Nupponen NN, Kakkola L, Koivisto P, Visakorpi T: **Genetic alterations in hormone-refractory recurrent prostate carcinomas.** *Am J Pathol* 1998, **153**:141-148.
18. Reiter RE, Sato I, Thomas G, Qian J, Gu Z, Watabe T, Loda M, Jenkins RB: **Coamplification of prostate stem cell antigen (PSCA) and MYC in locally advanced prostate cancer.** *Genes Chromosomes Cancer* 2000, **27**:95-103.
19. Watabe T, Lin M, Donjacour AA, Cunha GR, Witte ON, Reiter RE: **Growth, regeneration, and tumorigenesis of the prostate activates the PSCA promoter.** *Proc Natl Acad Sci USA* 2002, **99**:401-406.
20. Bahrenberg G, Brauers A, Joost HG, Jakse G: **PSCA expression is regulated by phorbol ester and cell adhesion in the bladder carcinoma cell line RT112.** *Cancer Lett* 2001, **168**:37-43.
21. Noda S, Kosugi A, Saitoh S, Narumiya S, Hamaoka T: **Protection from anti-TCR/CD3-induced apoptosis in immature thymocytes by a signal through thymic shared antigen-1/stem cell antigen-2.** *J Exp Med* 1996, **183**:2355-2360.
22. Thomas PM, Samelson LE: **The glycophosphatidylinositol-anchored Thy-1 molecule interacts with the p60fyn protein tyrosine kinase in T cells.** *J Biol Chem* 1992, **267**:12317-12322.
23. Bamezai A, Rock KL: **Overexpressed Ly-6A.2 mediated cell-cell adhesion by binding a ligand expressed on lymphoid cells.** *Proc Natl Acad Sci USA* 1995, **92**:4294-4298.
24. Katz BZ, Eshel R, Sagi-Assif O, Witz IP: **An association between high Ly-6A/E expression on tumor cells and a highly malignant phenotype.** *Int J Cancer* 1994, **59**:684-691.
25. Brakenhoff RH, Gerretsen M, Knippels EM, van Dijk M, van Essen H, Weghuis DO, Sinke RJ, Snow GB, van Dongen GA: **The human E48 antigen, highly homologous to the murine Ly-6 antigen ThB, is a GPI-anchored molecule apparently involved in keratinocyte cell-cell adhesion.** *J Cell Biol* 1995, **129**:1677-1689.

Review

Translation Initiation in Cancer: A Novel Target for Therapy[1]

Funda Meric[2] and Kelly K. Hunt

Department of Surgical Oncology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030

Abstract

Translation initiation is regulated in response to nutrient availability and mitogenic stimulation and is coupled with cell cycle progression and cell growth. Several alterations in translational control occur in cancer. Variant mRNA sequences can alter the translational efficiency of individual mRNA molecules, which in turn play a role in cancer biology. Changes in the expression or availability of components of the translational machinery and in the activation of translation through signal transduction pathways can lead to more global changes, such as an increase in the overall rate of protein synthesis and translational activation of the mRNA molecules involved in cell growth and proliferation. We review the basic principles of translational control, the alterations encountered in cancer, and selected therapies targeting translation initiation to help elucidate new therapeutic avenues.

Introduction

The fundamental principle of molecular therapeutics in cancer is to exploit the differences in gene expression between cancer cells and normal cells. With the advent of cDNA array technology, most efforts have concentrated on identifying differences in gene expression at the level of mRNA, which can be attributable either to DNA amplification or to differences in transcription. Gene expression is quite complicated, however, and is also regulated at the level of mRNA stability, mRNA translation, and protein stability.

The power of translational regulation has been best recognized among developmental biologists, because transcription does not occur in early embryogenesis in eukaryotes. For example, in *Xenopus*, the period of transcriptional quiescence continues until the embryo reaches midblastula transition, the 4000-cell stage. Therefore, all necessary mRNA molecules are transcribed during oogenesis and stockpiled in a translationally inactive, masked form. The mRNA are translationally activated at appropriate times during oocyte maturation, fertilization, and early embryogenesis and thus, are under strict translational control.

Translation has an established role in cell growth. Basically, an increase in protein synthesis occurs as a consequence of mitogenesis. Until recently, however, little was known about the alterations in mRNA translation in cancer, and much is yet to be discovered about their role in the development and progression of cancer. Here we review the basic principles of translational control, the alterations encountered in cancer, and selected therapies targeting translation initiation to elucidate potential new therapeutic avenues.

Basic Principles of Translational Control

Mechanism of Translation Initiation

Translation initiation is the main step in translational regulation. Translation initiation is a complex process in which the initiator tRNA and the 40S and 60S ribosomal subunits are recruited to the 5′ end of a mRNA molecule and assembled by eukaryotic translation initiation factors into an 80S ribosome at the start codon of the mRNA (Fig. 1). The 5′ end of eukaryotic mRNA is capped, *i.e.*, contains the cap structure m^7GpppN (7-methylguanosine-triphospho-5′-ribonucleoside). Most translation in eukaryotes occurs in a cap-dependent fashion, *i.e.*, the cap is specifically recognized by the eIF4E,[3] which binds the 5′ cap. The eIF4F translation initiation complex is then formed by the assembly of eIF4E, the RNA helicase eIF4A, and eIF4G, a scaffolding protein that mediates the binding of the 40S ribosomal subunit to the mRNA molecule through interaction with the eIF3 protein present on the 40S ribosome. eIF4A and eIF4B participate in melting the secondary structure of the 5′ UTR of the mRNA. The 43S initiation complex (40S/eIF2/Met-tRNA/GTP complex) scans the mRNA in a 5′→3′ direction until it encounters an AUG start codon. This start codon is then base-paired to the anticodon of initiator tRNA, forming the 48S initiation complex. The initiation factors are then displaced from the 48S complex, and the 60S ribosome joins to form the 80S ribosome.

Unlike most eukaryotic translation, translation initiation of certain mRNAs, such as the picornavirus RNA, is cap independent and occurs by internal ribosome entry. This mechanism does not require eIF4E. Either the 43S complex can bind the initiation codon directly through interaction with the IRES in the 5′ UTR such as in the encephalomyocarditis virus, or it can

Received 5/16/02; revised 7/12/02; accepted 7/22/02.

[1] F. M. is supported by The University of Texas M. D. Anderson Cancer Center Physician-Scientist Program and by NIH Grant 1KO8-CA 91895-01. K. K. H. is supported by Department of Defense Award DAMD-17-97-1-7162.

[2] To whom requests for reprints should be addressed, at Department of Surgical Oncology, Box 444, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030. Phone: (713) 745-4453; Fax: (713) 745-4926; E-mail: fmeric@mdanderson.org.

[3] The abbreviations used are: eIF4E, eukaryotic initiation factor 4E; UTR, untranslated region; IRES, internal ribosome entry site; 4E-BP1, eukaryotic initiation factor 4E-binding protein 1; S6K, ribosomal p70 S6 kinase; mTOR, mammalian target of rapamycin; ATM, ataxia telangiectasia mutated; PI3K, phosphatidylinositol 3-kinase; PTEN, phosphatase and tensin homolog deleted from chromosome 10; PP2A, protein phosphatase 2A; TGF-β3, transforming growth factor-β3; PAP, poly(A) polymerase; EPA, eicosopentaenoic acid; *mda-7*, melanoma differentiation-associated gene 7.

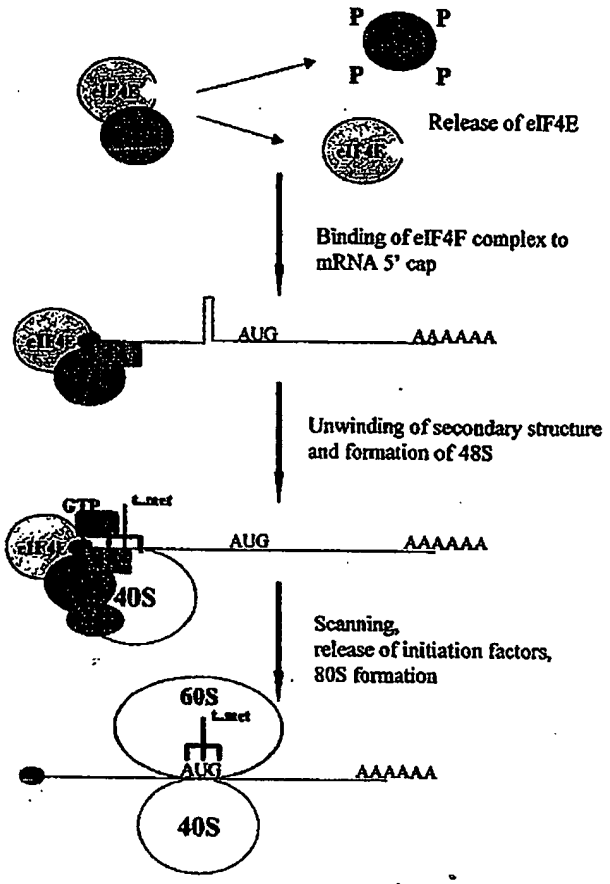


Fig. 1. Translation initiation in eukaryotes. The 4E-BPs are hyperphosphorylated to release eIF4E so that it can interact with the 5′ cap, and the eIF4F initiation complex is assembled. The interaction of poly(A) binding protein with the initiation complex and circularization of the mRNA is not depicted in the diagram. The secondary structure of the 5′ UTR is melted, the 40S ribosomal subunit is bound td eIF3, and the ternary complex consisting of eIF2, GTP, and the Met-tRNA are recruited to the mRNA. The ribosome scans the mRNA in a 5′→3′ direction until an AUG start codon is found in the appropriate sequence context. The initiation factors are released, and the large ribosomal subunit is recruited.

initially attach to the IRES and then reach the initiation codon by scanning or transfer, as is the case with the poliovirus (1).

Regulation of Translation Initiation

Translation initiation can be regulated by alterations in the expression or phosphorylation status of the various factors involved. Key components in translational regulation that may provide potential therapeutic targets follow.

eIF4E. eIF4E plays a central role in translation regulation. It is the least abundant of the initiation factors and is considered the rate-limiting component for initiation of cap-dependent translation. eIF4E may also be involved in mRNA splicing, mRNA 3′ processing, and mRNA nucleocytoplasmic transport (2). eIF4E expression can be increased at the transcriptional level in response to serum or growth factors (3). eIF4E overexpression may cause preferential translation of mRNAs containing excessive secondary structure in their 5′ UTR that are normally discriminated against by the translational machinery and thus are inefficiently translated (4–7). As examples of this, overexpression of eIF4E promotes increased translation of vascular endothelial growth factor, fibroblast growth factor-2, and cyclin D1 (2, 8, 9).

Another mechanism of control is the regulation of eIF4E phosphorylation. eIF4E phosphorylation is mediated by the mitogen-activated protein kinase-interacting kinase 1, which is activated by the mitogen-activated pathway activating extracellular signal-related kinases and the stress-activated pathway acting through p38 mitogen-activated protein kinase (10–13). Several mitogens, such as serum, platelet-derived growth factor, epidermal growth factor, insulin, angiotensin II, src kinase overexpression, and ras overexpression, lead to eIF4E phosphorylation (14). The phosphorylation status of eIF4E is usually correlated with the translational rate and growth status of the cell; however; eIF4E phosphorylation has also been observed in response to some cellular stresses when translational rates actually decrease (15). Thus, further study is needed to understand the effects of eIF4E phosphorylation on eIF4E activity.

Another mechanism of regulation is the alteration of eIF4E availability by the binding of eIF4E to the eIF4E-binding proteins (4E-BP, also known as PHAS-I). 4E-BPs compete with eIF4G for a binding site in eIF4E. The binding of eIF4E to the best characterized eIF4E-binding protein, 4E-BP1, is regulated by 4E-BP1 phosphorylation. Hypophosphorylated 4E-BP1 binds to eIF4E, whereas 4E-BP1 hyperphosphorylation decreases this binding. Insulin, angiotensin, epidermal growth factor, platelet-derived growth factor, hepatocyte growth factor, nerve growth factor, insulin-like growth factors I and II, interleukin 3, granulocyte-macrophage colony-stimulating factor + steel factor, gastrin, and the adenovirus have all been reported to induce phosphorylation of 4E-BP1 and to decrease the ability of 4E-BP1 to bind eIF4E (15, 16). Conversely, deprivation of nutrients or growth factors results in 4E-BP1 dephosphorylation, an increase in eIF4E binding, and a decrease in cap-dependent translation.

p70 S6 Kinase. Phosphorylation of ribosomal 40S protein S6 by S6K is thought to play an important role in translational regulation. S6K −/− mouse embryonic cells proliferate more slowly than do parental cells, demonstrating that S6K has a positive influence on cell proliferation (17). S6K regulates the translation of a group of mRNAs possessing a 5′ terminal oligopyrimidine tract (5′ TOP) found at the 5′ UTR of ribosomal protein mRNAs and other mRNAs coding for components of the translational machinery. Phosphorylation of S6K is regulated in part based on the availability of nutrients (18, 19) and is stimulated by several growth factors, such as platelet-derived growth factor and insulin-like growth factor I (20).

eIF2α Phosphorylation. The binding of the initiator tRNA to the small ribosomal unit is mediated by translation initiation factor eIF2. Phosphorylation of the α-subunit of eIF2 prevents formation of the eIF2/GTP/Met-tRNA complex and inhibits global protein synthesis (21, 22). eIF2α is phosphorylated under a variety of conditions, such as viral infection, nutrient deprivation, heme deprivation, and apoptosis (22). eIF2α is phosphorylated by heme-regulated inhibitor, nutrient-regulated protein kinase, and the IFN-induced, double-stranded RNA-activated protein kinase (PKR; Ref. 23).

The mTOR Signaling Pathway. The macrolide antibiotic rapamycin (Sirolimus; Wyeth-Ayerst Research, Collegeville, PA) has been the subject of intensive study because it inhibits signal transduction pathways involved in T-cell activation. The rapamycin-sensitive component of these pathways is mTOR (also called FRAP or RAFT1). mTOR is the mammalian homologue of the yeast TOR proteins that regulate G_1 progression and translation in response to nutrient availability (24). mTOR is a serine-threonine kinase that modulates translation initiation by altering the phosphorylation status of 4E-BP1 and S6K (Fig. 2; Ref. 25).

4E-BP1 is phosphorylated on multiple residues. mTOR phosphorylates the Thr-37 and Thr-46 residues of 4E-BP1 *in vitro* (26); however, phosphorylation at these sites is not associated with a loss of eIF4E binding. Phosphorylation of Thr-37 and Thr-46 is required for subsequent phosphorylation at several COOH-terminal, serum-sensitive sites; a combination of these phosphorylation events appears to be needed to inhibit the binding of 4E-BP1 to eIF4E (25). The product of the *ATM* gene, p38/MSK1 pathway, and protein kinase Cσ also play a role in 4E-BP1 phosphorylation (27–29).

S6K and 4E-BP1 are also regulated, in part, by PI3K and its downstream protein kinase Akt. PTEN is a phosphatase that negatively regulates PI3K signaling. PTEN null cells have constitutively active of Akt, with increased S6K activity and S6 phosphorylation (30). S6K activity is inhibited both by PI3K inhibitors wortmannin and LY294002 and by mTOR inhibitor rapamycin (24). Akt phosphorylates Ser-2448 in mTOR *in vitro*, and this site is phosphorylated upon Akt activation *in vivo* (31–33). Thus, mTOR is regulated by the PI3K/Akt pathway; however, this does not appear to be the only mode of regulation of mTOR activity. Whether the PI3K pathway also regulates S6K and 4E-BP1 phosphorylation independent of mTOR is controversial.

Interestingly, mTOR autophosphorylation is blocked by wortmannin but not by rapamycin (34). This seeming inconsistency suggests that mTOR-responsive regulation of 4E-BP1 and S6K activity occurs through a mechanism other than intrinsic mTOR kinase activity. An alternate pathway for 4E-BP1 and S6K phosphorylation by mTOR activity is by the inhibition of a phosphatase. Treatment with calyculin A, an inhibitor of phosphatases 1 and 2A, reduces rapamycin-induced dephosphorylation of 4E-BP1 and S6K by rapamycin (35). PP2A interacts with full-length S6K but not with a S6K mutant that is resistant to dephosphorylation resulting from rapamycin. mTOR phosphorylates PP2A *in vitro*; however, how this process alters PP2A activity is not known. These results are consistent with the model that phosphorylation of a phosphatase by mTOR prevents dephosphorylation of 4E-BP1 and S6K, and conversely, that nutrient deprivation and rapamycin block inhibition of the phosphatase by mTOR.

Polyadenylation. The poly(A) tail in eukaryotic mRNA is important in enhancing translation initiation and mRNA stability. Polyadenylation plays a key role in regulating gene expression during oogenesis and early embryogenesis. Some mRNA that are translationally inactive in the oocyte are polyadenylated concomitantly with translational activation in oocyte maturation, whereas other mRNAs that are translationally active during oogenesis are deadenylated and translationally silenced (36–38). Thus, control of poly(A) tail synthesis is an important regulatory step in gene expression. The 5′ cap and poly(A) tail are thought to function synergistically to regulate mRNA translational efficiency (39, 40).

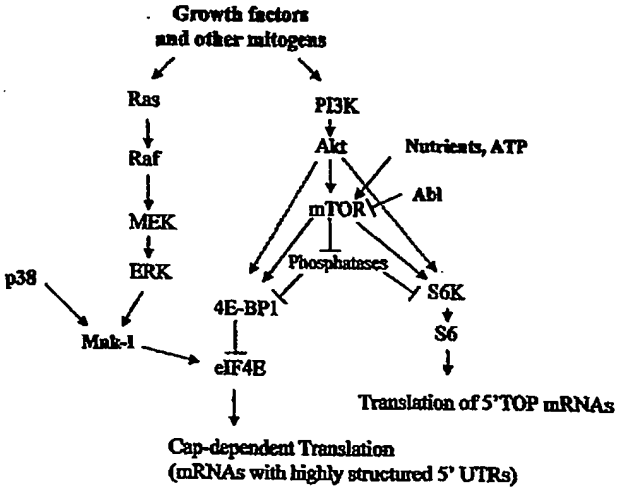


Fig. 2. Regulation of translation initiation by signal transduction pathways. Signaling via p38, extracellular signal-related kinase, PI3K, and mTOR can all activate translation initiation.

RNA Packaging. Most RNA-binding proteins are assembled on a transcript at the time of transcription, thus determining the translational fate of the transcript (41). A highly conserved family of Y-box proteins is found in cytoplasmic messenger ribonucleoprotein particles, where the proteins are thought to play a role in restricting the recruitment of mRNA to the translational machinery (41–43). The major mRNA-associated protein, YB-1, destabilizes the interaction of eIF4E and the 5′ mRNA cap *in vitro*, and overexpression of YB-1 results in translational repression *in vivo* (44). Thus, alterations in RNA packaging can also play an important role in translational regulation.

Translation Alterations Encountered in Cancer

Three main alterations at the translational level occur in cancer: variations in mRNA sequences that increase or decrease translational efficiency, changes in the expression or availability of components of the translational machinery, and activation of translation through aberrantly activated signal transduction pathways. The first alteration affects the translation of an individual mRNA that may play a role in carcinogenesis. The second and third alterations can lead to more global changes, such as an increase in the overall rate of protein synthesis, and the translational activation of several mRNA species.

Variations in mRNA Sequence

Variations in mRNA sequence affect the translational efficiency of the transcript. A brief description of these variations and examples of each mechanism follow.

Mutations. Mutations in the mRNA sequence, especially in the 5′ UTR, can alter its translational efficiency, as seen in the following examples.

c-myc. Saito *et al.* proposed that translation of full-length *c-myc* is repressed, whereas in several Burkitt lymphomas that have deletions of the mRNA 5′ UTR, translation of *c-myc* is more efficient (45). More recently, it was reported that the 5′ UTR of *c-myc* contains an IRES, and thus *c-myc* translation can be initiated by a cap-independent as well as a cap-dependent mechanism (46, 47). In patients with multiple myeloma, a C→T mutation in the c-myc IRES was identified (48) and found to cause an enhanced initiation of translation via internal ribosomal entry (49).

BRCA1. A somatic point mutation (117 G→C) in position −3 with respect to the start codon of the *BRCA1* gene was identified in a highly aggressive sporadic breast cancer (50). Chimeric constructs consisting of the wild-type or mutated *BRCA1* 5′ UTR and a downstream luciferase reporter demonstrated a decrease in the translational efficiency with the 5′ UTR mutation.

Cyclin-dependent Kinase Inhibitor 2A. Some inherited melanoma kindreds have a G→T transversion at base −34 of cyclin-dependent kinase inhibitor-2A, which encodes a cyclin-dependent kinase 4/cyclin-dependent kinase 6 kinase inhibitor important in G_1 checkpoint regulation (51). This mutation gives rise to a novel AUG translation initiation codon, creating an upstream open reading frame that competes for scanning ribosomes and decreases translation from the wild-type AUG.

Alternate Splicing and Alternate Transcription Start Sites. Alterations in splicing and alternate transcription sites can lead to variations in 5′ UTR sequence, length, and secondary structure, ultimately impacting translational efficiency.

ATM. The *ATM* gene has four noncoding exons in its 5′ UTR that undergo extensive alternative splicing (52). The contents of 12 different 5′ UTRs that show considerable diversity in length and sequence have been identified. These divergent 5′ leader sequences play an important role in the translational regulation of the *ATM* gene.

mdm. In a subset of tumors, overexpression of the oncoprotein mdm2 results in enhanced translation of the *mdm2* mRNA. Use of different promoters leads to two *mdm2* transcripts that differ only in their 5′ leaders (53). The longer 5′ UTR contains two upstream open reading frames, and this mRNA is loaded with ribosomes inefficiently compared with the short 5′ UTR.

BRCA1. In a normal mammary gland, *BRCA1* mRNA is expressed with a shorter leader sequence (5′UTRa), whereas in sporadic breast cancer tissue, *BRCA1* mRNA is expressed with a longer leader sequence (5′ UTRb); the translational efficiency of transcripts containing 5′ UTRb is 10 times lower than that of transcripts containing 5′ UTRa (54).

TGF-β3. *TGF-β3* mRNA includes a 1.1-kb 5′ UTR, which exerts an inhibitory effect on translation. Many human breast cancer cell lines contain a novel *TGF-β3* transcript with a 5′ UTR that is 870 nucleotides shorter and has a 7-fold greater translational efficiency than the normal *TGF-β3* mRNA (55).

Alternate Polyadenylation Sites. Multiple polyadenylation signals leading to the generation of several transcripts with differing 3′ UTR have been described for several mRNA species, such as the *RET* proto-oncogene (56), *ATM* gene (52), tissue inhibitor of metalloproteinases-3 (57), *RHOA* proto-oncogene (58), and calmodulin-I (59). Although the effect of these alternate 3′ UTRs on translation is not yet known, they may be important in RNA-protein interactions that affect translational recruitment. The role of these alterations in cancer development and progression is unknown.

Alterations in the Components of the Translation Machinery

Alterations in the components of translation machinery can take many forms.

Overexpresssion of eIF4E. Overexpression of eIF4E causes malignant transformation in rodent cells (60) and the deregulation of HeLa cell growth (61). Polunovsky *et al.* (62) found that eIF4E overexpression substitutes for serum and individual growth factors in preserving viability of fibroblasts, which suggests that eIF4E can mediate both proliferative and survival signaling.

Elevated levels of eIF4E mRNA have been found in a broad spectrum of transformed cell lines (63). eIF4E levels are elevated in all ductal carcinoma *in situ* specimens and invasive ductal carcinomas, compared with benign breast specimens evaluated with Western blot analysis (64, 65). Preliminary studies suggest that this overexpression is attributable to gene amplification (66).

There are accumulating data suggesting that eIF4E overexpression can be valuable as a prognostic marker. eIF4E overexpression was found in a retrospective study to be a marker of poor prognosis in stages I to III breast carcinoma (67). Verification of the prognostic value of eIF4E in breast cancer is now under way in a prospective trial (67). However, in a different study, eIF4E expression was correlated with the aggressive behavior of non-Hodgkin's lymphomas (68). In a prospective analysis of patients with head and neck cancer, elevated levels of eIF4E in histologically tumor-free surgical margins predicted a significantly increased risk of local-regional recurrence (9). These results all suggest that eIF4E overexpression can be used to select patients who might benefit from more aggressive systemic therapy. Furthermore, the head and neck cancer data suggest that eIF4E overexpression is a field defect and can be used to guide local therapy.

Alterations in Other Initiation Factors. Alterations in a number of other initiation factors have been associated with cancer. Overproduction of eIF4G, similar to eIF4E, leads to malignant transformation *in vitro* (69). eIF-2α is found in increased levels in bronchioloalveolar carcinomas of the lung (3). Initiation factor eIF-4A1 is overexpressed in melanoma (70) and hepatocellular carcinoma (71). The p40 subunit of translation initiation factor 3 is amplified and overexpressed in breast and prostate cancer (72), and the eIF3-p110 subunit is overexpressed in testicular seminoma (73). The role that overexpression of these initiation factors plays on the development and progression of cancer, if any, is not known.

Overexpression of S6K. S6K is amplified and highly overexpressed in the MCF7 breast cancer cell line, compared with normal mammary epithelium (74). In a study by Barlund *et al.* (74), S6K was amplified in 59 of 668 primary breast tumors, and a statistically significant association was observed between amplification and poor prognosis.

Overexpression of PAP. PAP catalyzes 3′ poly(A) synthesis. PAP is overexpressed in human cancer cells compared with normal and virally transformed cells (75). PAP enzymatic activity in breast tumors has been correlated with PAP protein levels (76) and, in mammary tumor cytosols, was found to be an independent factor for predicting survival (76). Little is known, however, about how PAP expression or activity affects the translational profile.

Alterations in RNA-binding Proteins. Even less is known about alterations in RNA packaging in cancer. Increased expression and nuclear localization of the RNA-binding protein YB-1 are indicators of a poor prognosis for breast cancer (77), non-small cell lung cancer (78), and ovarian cancer (79). However, this effect may be mediated at least in part at the level of transcription, because YB-1 increases chemoresistance by enhancing the transcription of a multidrug resistance gene (80).

Activation of Signal Transduction Pathways

Activation of signal transduction pathways by loss of tumor suppressor genes or overexpression of certain tyrosine kinases can contribute to the growth and aggressiveness of tumors. An important mutant in human cancers is the tumor suppressor gene *PTEN*, which leads to the activation of the PI3K/Akt pathway. Activation of PI3K and Akt induces the oncogenic transformation of chicken embryo fibroblasts. The transformed cells show constitutive phosphorylation of S6K and of 4E-BP1 (81). A mutant Akt that retains kinase activity but does not phosphorylate S6K or 4E-BP1 does not transform fibroblasts, which suggests a correlation between the oncogenicity of PI3K and Akt and the phosphorylation of S6K and 4E-BP1 (81).

Several tyrosine kinases such as platelet-derived growth factor, insulin-like growth factor, HER2/neu, and epidermal growth factor receptor are overexpressed in cancer. Because these kinases activate downstream signal transduction pathways known to alter translation initiation, activation of translation is likely to contribute to the growth and aggressiveness of these tumors. Furthermore, the mRNA for many of these kinases themselves are under translational control. For example, HER2/neu mRNA is translationally controlled both by a short upstream open reading frame that represses HER2/neu translation in a cell type-independent manner and by a distinct cell type-dependent mechanism that increases translational efficiency (82). HER2/neu translation is different in transformed and normal cells. Thus, it is possible that alterations at the translational level can in part account for the discrepancy between *HER2/neu* gene amplification detected by fluorescence *in situ* hybridization and protein levels detected by immunohistochemical assays.

Translation Targets of Selected Cancer Therapy

Components of the translation machinery and signal pathways involved in the activation of translation initiation represent good targets for cancer therapy.

Targeting the mTOR Signaling Pathway: Rapamycin and Tumstatin

Rapamycin inhibits the proliferation of lymphocytes. It was initially developed as an immunosuppressive drug for organ transplantation. Rapamycin with FKBP 12 (FK506-binding protein, M_r 12,000) binds to mTOR to inhibit its function.

Rapamycin causes a small but significant reduction in the initiation rate of protein synthesis (83). It blocks cell growth in part by blocking S6 phosphorylation and selectively suppressing the translation of 5′ TOP mRNAs, such as ribosomal proteins, and elongation factors (83–85). Rapamycin also blocks 4E-BP1 phosphorylation and inhibits cap-dependent but not cap-independent translation (17, 86).

The rapamycin-sensitive signal transduction pathway, activated during malignant transformation and cancer progression, is now being studied as a target for cancer therapy (87). Prostate, breast, small cell lung, glioblastoma, melanoma, and T-cell leukemia are among the cancer lines most sensitive to the rapamycin analogue CCI-779 (Wyeth-Ayerst Research; Ref. 87). In rhabdomyosarcoma cell lines, rapamycin is either cytostatic or cytocidal, depending on the p53 status of the cell; p53 wild-type cells treated with rapamycin arrest in the G_1 phase and maintain their viability, whereas p53 mutant cells accumulate in G_1 and undergo apoptosis (88, 89). In a recently reported study using human primitive neuroectodermal tumor and medulloblastoma models, rapamycin exhibited more cytotoxicity in combination with cisplatin and camptothecin than as a single agent. *In vivo*, CCI-779 delayed growth of xenografts by 160% after 1 week of therapy and 240% after 2 weeks. A single high-dose administration caused a 37% decrease in tumor volume. Growth inhibition *in vivo* was 1.3 times greater, with cisplatin in combination with CCI-779 than with cisplatin alone (90). Thus, preclinical studies suggest that rapamycin analogues are useful as single agents and in combination with chemotherapy.

Rapamycin analogues CCI-779 and RAD001 (Novartis, Basel, Switzerland) are now in clinical trials. Because of the known effect of rapamycin on lymphocyte proliferation, a potential problem with rapamycin analogues is immunosuppression. However, although prolonged immunosuppression can result from rapamycin and CCI-779 administered on continuous-dose schedules, the immunosuppressive effects of rapamycin analogues resolve in ~24 h after therapy (91). The principal toxicities of CCI-779 have included dermatological toxicity, myelosuppression, infection, mucositis, diarrhea, reversible elevations in liver function tests, hyperglycemia, hypokalemia, hypocalcemia, and depression (87, 92–94). Phase II trials of CCI-779 have been conducted in advanced renal cell carcinoma and in stage III/IV breast carcinoma patients who failed with prior chemotherapy. In the results reported in abstract form, although there were no complete responses, partial responses were documented in both renal cell carcinoma and in breast carcinoma (94, 95). Thus, CCI-779 has documented preliminary clinical activity in a previously treated, unselected patient population.

Active investigation is under way into patient selection for mTOR inhibitors. Several studies have found an enhanced efficacy of CCI-779 in PTEN-null tumors (30, 96). Another study found that six of eight breast cancer cell lines were responsive to CCI-779, although only two of these lines lacked PTEN (97) There was, however, a positive correlation between Akt activation and CCI-779 sensitivity (97). This correlation suggests that activation of the PI3K-Akt pathway,

regardless of whether it is attributable to a PTEN mutation or to overexpression of receptor tyrosine kinases, makes cancer cell amenable to mTOR-directed therapy. In contrast, lower levels of the target of mTOR, 4E-BP1, are associated with rapamycin resistance; thus, a lower 4E-BP1/eIF4E ratio may predict rapamycin resistance (98).

Another mode of activity for rapamycin and its analogues appears to be through inhibition of angiogenesis. This activity may be both through direct inhibition of endothelial cell proliferation as a result of mTOR inhibition in these cells or by inhibition of translation of such proangiogenic factors as vascular endothelial growth factor in tumor cells (99, 100).

The angiogenesis inhibitor tumstatin, another anticancer drug currently under study, was also found recently to inhibit translation in endothelial cells (101). Through a requisite interaction with integrin, tumstatin inhibits activation of the PI3K/Akt pathway and mTOR in endothelial cells and prevents dissociation of eIF4E from 4E-BP1, thereby inhibiting cap-dependent translation. These findings suggest that endothelial cells are especially sensitive to therapies targeting the mTOR-signaling pathway.

Targeting eIF2α: EPA, Clotrimazole, mda-7, and Flavonoids

EPA is an n-3 polyunsaturated fatty acid found in the fish-based diets of populations having a low incidence of cancer (102). EPA inhibits the proliferation of cancer cells (103), as well as in animal models (104, 105). It blocks cell division by inhibiting translation initiation (105). EPA releases Ca^{2+} from intracellular stores while inhibiting their refilling, thereby activating PKR. PKR, in turn phosphorylates and inhibits eIF2α, resulting in the inhibition of protein synthesis at the level of translation initiation. Similarly, clotrimazole, a potent antiproliferative agent *in vitro* and *in vivo*, inhibits cell growth through depletion of Ca^{2+} stores, activation of PKR, and phosphorylation of eIF2α (106). Consequently, clotrimazole preferentially decreases the expression of cyclins A, E, and D1, resulting in blockage of the cell cycle in G_1.

mda-7 is a novel tumor suppressor gene being developed as a gene therapy agent. Adenoviral transfer of *mda-7* (Ad-mda7) induces apoptosis in many cancer cells including breast, colorectal, and lung cancer (107–109). Ad-mda7 also induces and activates PKR, which leads to phosphorylation of eIF2α and induction of apoptosis (110).

Flavonoids such as genistein and quercetin suppress tumor cell growth. All three mammalian eIF2α kinases, PKR, heme-regulated inhibitor, and PERK/PEK, are activated by flavonoids, with phosphorylation of eIF2α and inhibition of protein synthesis (111).

Targeting eIF4A and eIF4E: Antisense RNA and Peptides

Antisense expression of eIF4A decreases the proliferation rate of melanoma cells (112). Sequestration of eIF4E by overexpression of 4E-BP1 is proapoptotic and decreases tumorigenicity (113, 114). Reduction of eIF4E with antisense RNA decreases soft agar growth, increases tumor latency, and increases the rates of tumor doubling times (7). Antisense eIF4E RNA treatment also reduces the expression of angiogenic factors (115) and has been proposed as a potential adjuvant therapy for head and neck cancers, particularly when elevated eIF4E is found in surgical margins. Small molecule inhibitors that bind the eIF4G/4E-BP1-binding domain of eIF4E are proapoptotic (116) and are also being actively pursued.

Exploiting Selective Translation for Gene Therapy

A different therapeutic approach that takes advantage of the enhanced cap-dependent translation in cancer cells is the use of gene therapy vectors encoding suicide genes with highly structured 5′ UTR. These mRNA would thus be at a competitive disadvantage in normal cells and not translate well, whereas in cancer cells, they would translate more efficiently. For example, the introduction of the 5′ UTR of fibroblast growth factor-2 5′ to the coding sequence of *herpes simplex virus type-1 thymidine kinase* gene, allows for selective translation of *herpes simplex virus type-1 thymidine kinase* gene in breast cancer cell lines compared with normal mammary cell lines and results in selective sensitivity to ganciclovir (117).

Toward the Future

Translation is a crucial process in every cell. However, several alterations in translational control occur in cancer. Cancer cells appear to need an aberrantly activated translational state for survival, thus allowing the targeting of translation initiation with surprisingly low toxicity. Components of the translational machinery, such as eIF4E, and signal transduction pathways involved in translation initiation, such mTOR, represent promising targets for cancer therapy. Inhibitors of the mTOR have already shown some preliminary activity in clinical trials. It is possible that with the development of better predictive markers and better patient selection, response rates to single-agent therapy can be improved. Similar to other cytostatic agents, however, mTOR inhibitors are most likely to achieve clinical utility in combination therapy. In the interim, our increasing understanding of translation initiation and signal transduction pathways promise to lead to the identification of new therapeutic targets in the near future.

Acknowledgments

We thank Gayle Nesom from The University of Texas M. D. Anderson Cancer Center Department of Scientific Publications for editorial assistance and Dr. Elmer Bernstam for assistance with manuscript preparation.

References

1. Pestova, T. V., Kolupaeva, V. G., Lomakin, I. B., Pilipenko, E. V., Shatsky, I. N., Agol, V. I., and Hellen, C. U. Molecular mechanisms of translation initiation in eukaryotes. Proc. Natl. Acad. Sci. USA, *98:* 7029–7036, 2001.

2. Rosenwald, I. B., Kaspar, R., Rousseau, D., Gehrke, L., Leboulch, P., Chen, J. J., Schmidt, E. V., Sonenberg, N., and London, I. M. Eukaryotic translation initiation factor 4E regulates expression of cyclin D1 at transcriptional and post-transcriptional levels. J. Biol. Chem., *270:* 21176–21180, 1995.

3. Rosenwald, I. B., Hutzler, M. J., Wang, S., Savas, L., and Fraire, A. E. Expression of eukaryotic translation initiation factors 4E and 2α is increased frequently in bronchioloalveolar but not in squamous cell carcinomas of the lung. Cancer (Phila.), *92:* 2164–2171, 2001.

4. Darveau, A., Pelletier, J., and Sonenberg, N. Differential efficiencies of *in vitro* translation of mouse c-myc transcripts differing in the 5' untranslated region. Proc. Natl. Acad. Sci. USA, *82:* 2315–2319, 1985.

5. Kozak, M. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. Proc. Natl. Acad. Sci. USA, *83:* 2850–2854, 1986.

6. Koromilas, A. E., Lazaris-Karatzas, A., and Sonenberg, N. mRNAs containing extensive secondary structure in their 5' non-coding region translate efficiently in cells overexpressing initiation factor eIF-4E. EMBO J., *11:* 4153–4158, 1992.

7. Rinker-Schaeffer, C. W., Graff, J. R., De Benedetti, A., Zimmer, S. G., and Rhoads, R. E. Decreasing the level of translation initiation factor 4E with antisense RNA causes reversal of ras-mediated transformation and tumorigenesis of cloned rat embryo fibroblasts. Int. J. Cancer, *55:* 841–847, 1993.

8. Kevil, C. G., De Benedetti, A., Payne, D. K., Coe, L. L., Laroux, F. S., and Alexander, J. S. Translational regulation of vascular permeability factor by eukaryotic initiation factor 4E: implications for tumor angiogenesis. Int. J. Cancer, *65:* 785–790, 1996.

9. Nathan, C. A., Franklin, S., Abreo, F. W., Nassar, R., De Benedetti, A., and Glass, J. Analysis of surgical margins with the molecular marker eIF4E: a prognostic factor in patients with head and neck cancer. J. Clin. Oncol., *17:* 2909–2914, 1999.

10. Fukunaga, R., and Hunter, T. MNK1, a new MAP kinase-activated protein kinase, isolated by a novel expression screening method for identifying protein kinase substrates. EMBO J., *16:* 1921–1933, 1997.

11. Waskiewicz, A. J., Flynn, A., Proud, C. G., and Cooper, J. A. Mitogen-activated protein kinases activate the serine/threonine kinases Mnk1 and Mnk2. EMBO J., *16:* 1909–1920, 1997.

12. Wang, X., Flynn, A., Waskiewicz, A. J., Webb, B. L., Vries, R. G., Baines, I. A., Cooper, J. A., and Proud, C. G. The phosphorylation of eukaryotic initiation factor eIF4E in response to phorbol esters, cell stresses, and cytokines is mediated by distinct MAP kinase pathways. J. Biol. Chem., *273:* 9373–9377, 1998.

13. Pyronnet, S., Imataka, H., Gingras, A. C., Fukunaga, R., Hunter, T., and Sonenberg, N. Human eukaryotic translation initiation factor 4G (eIF4G) recruits Mnk1 to phosphorylate eIF4E. EMBO J., *18:* 270–279, 1999.

14. Kleijn, M., Scheper, G. C., Voorma, H. O., and Thomas, A. A. Regulation of translation initiation factors by signal transduction. Eur. J. Biochem., *253:* 531–544, 1998.

15. Raught, B., and Gingras, A. C. eIF4E activity is regulated at multiple levels. Int. J. Biochem. Cell Biol., *31:* 43–57, 1999.

16. Takeuchi, K., Shibamoto, S., Nagamine, K., Shigemori, I., Omura, S., Kitamura, N., and Ito, F. Signaling pathways leading to transcription and translation cooperatively regulate the transient increase in expression of c-Fos protein. J. Biol. Chem., *276:* 26077–26083, 2001.

17. Kawasome, H., Papst, P., Webb, S., Keller, G. M., Johnson, G. L., Gelfand, E. W., and Terada, N. Targeted disruption of p70(s6k) defines its role in protein synthesis and rapamycin sensitivity. Proc. Natl. Acad. Sci. USA, *95:* 5033–5038, 1998.

18. Christie, G. R., Hajduch, E., Hundal, H. S., Proud, C. G., and Taylor, P. M. Intracellular sensing of amino acids in *Xenopus laevis* oocytes stimulates p70 S6 kinase in a target of rapamycin-dependent manner. J. Biol. Chem., *277:* 9952–9957, 2002.

19. Hara, K., Yonezawa, K., Weng, Q. P., Kozlowski, M. T., Belham, C., and Avruch, J. Amino acid sufficiency and mTOR regulate p70 S6 kinase and eIF-4E BP1 through a common effector mechanism. J. Biol. Chem., *273:* 14484–14494, 1998.

20. Graves, L. M., Bornfeldt, K. E., Argast, G. M., Krebs, E. G., Kong, X., Lin, T. A., and Lawrence, J. C., Jr. cAMP- and rapamycin-sensitive regulation of the association of eukaryotic initiation factor 4E and the translational regulator PHAS-I in aortic smooth muscle cells. Proc. Natl. Acad. Sci. USA, *92:* 7222–7226, 1995.

21. Merrick, W. C., and Hershey, J. W. B. The pathway and mechanism of eukaryotic protein synthesis. *In:* J. W. B. Hershey and M. B. Mathews (eds.), Translational Control, pp. 31–69. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1996.

22. Kimball, S. R. Eukaryotic initiation factor eIF2. Int. J. Biochem. Cell Biol., *31:* 25–29, 1999.

23. Jagus, R., Joshi, B., and Barber, G. N. PKR, apoptosis and cancer. Int. J. Biochem. Cell Biol., *31:* 123–138, 1999.

24. Thomas, G., and Hall, M. N. TOR signalling and control of cell growth. Curr. Opin. Cell Biol., *9:* 782–787, 1997.

25. Gingras, A. C., Raught, B., and Sonenberg, N. Regulation of translation initiation by FRAP/mTOR. Genes Dev., *15:* 807–826, 2001.

26. Gingras, A. C., Gygi, S. P., Raught, B., Polakiewicz, R. D., Abraham, R. T., Hoekstra, M. F., Aebersold, R., and Sonenberg, N. Regulation of 4E-BP1 phosphorylation: a novel two-step mechanism. Genes Dev., *13:* 1422–1437, 1999.

27. Kumar, V., Pandey, P., Sabatini, D., Kumar, M., Majumder, P. K., Bharti, A., Carmichael, G., Kufe, D., and Kharbhanda, S. Functional interaction between RAFT1/FRAP/mTOR and protein kinase Cδ in the regulation of cap-dependent initiation of translation. EMBO J., *19:* 1087–1097, 2000.

28. Yang, D. Q., and Kastan, M. B. Participation of ATM in insulin signalling through phosphorylation of eIF-4E-binding protein 1. Nat. Cell. Biol., *2:* 893–898, 2000.

29. Liu, G., Zhang, Y., Bode, A. M., Ma, W. Y., and Dong, Z. Phosphorylation of 4E-BP1 is mediated by the p38/MSK1 pathway in response to UVB irradiation. J. Biol. Chem., *277:* 8810–8816, 2002.

30. Neshat, M. S., Mellinghoff, I. K., Tran, C., Stiles, B., Thomas, G., Petersen, R., Frost, P., Gibbons, J. J., Wu, H., and Sawyers, C. L. Enhanced sensitivity of PTEN-deficient tumors to inhibition of FRAP/mTOR. Proc. Natl. Acad. Sci. USA, *98:* 10314–10319, 2001.

31. Sekulic, A., Hudson, C. C., Homme, J. L., Yin, P., Otterness, D. M., Karnitz, L. M., and Abraham, R. T. A direct linkage between the phosphoinositide 3-kinase-AKT signaling pathway and the mammalian target of rapamycin in mitogen-stimulated and transformed cells. Cancer Res., *60:* 3504–3513, 2000.

32. Scott, P. H., and Lawrence, J. C., Jr. Attenuation of mammalian target of rapamycin activity by increased cAMP in 3T3-L1 adipocytes. J. Biol. Chem., *273:* 34496–34501, 1998.

33. Reynolds, T. T., Bodine, S. C., and Lawrence, J. C., Jr. Control of Ser2448 phosphorylation in the mammalian target of rapamycin by insulin and skeletal muscle load. J. Biol. Chem., *277:* 17657–17662, 2002.

34. Peterson, R. T., Beal, P. A., Comb, M. J., and Schreiber, S. L. FKBP12-rapamycin-associated protein (FRAP) autophosphorylates at serine 2481 under translationally repressive conditions. J. Biol. Chem., *275:* 7416–7423, 2000.

35. Peterson, R. T., Desai, B. N., Hardwick, J. S., and Schreiber, S. L. Protein phosphatase 2A interacts with the 70-kDa S6 kinase and is activated by inhibition of FKBP12-rapamycin-associated protein. Proc. Natl. Acad. Sci. USA, *96:* 4438–4442, 1999.

36. McGrew, L. L., Dworkin-Rastl, E., Dworkin, M. B., and Richter, J. D. Poly(A) elongation during *Xenopus* oocyte maturation is required for translational recruitment and is mediated by a short sequence element. Genes Dev., *3:* 803–815, 1989.

37. Sheets, M. D., Wu, M., and Wickens, M. Polyadenylation of c-mos mRNA as a control point in *Xenopus* meiotic maturation. Nature (Lond.), *374:* 511–516, 1995.

38. Varnum, S. M., and Wormington, W. M. Deadenylation of maternal mRNAs during *Xenopus* oocyte maturation does not require specific *cis*-sequences: a default mechanism for translational control. Genes Dev., *4:* 2278–2286, 1990.

39. Gallie, D. R. The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. Genes Dev., *5:* 2108–2116, 1991.

40. Sachs, A. B., and Varani, G. Eukaryotic translation initiation: there are (at least) two sides to every story. Nat. Struct. Biol., *7:* 356–361, 2000.

41. Wolffe, A. P., and Meric, F. Coupling transcription to translation: a novel site for the regulation of eukaryotic gene expression. Int. J. Biochem. Cell Biol., *28:* 247–257, 1996.

42. Evdokimova, V. M., Wei, C. L., Sitikov, A. S., Simonenko, P. N., Lazarev, O. A., Vasilenko, K. S., Ustinov, V. A., Hershey, J. W., and Ovchinnikov, L. P. The major protein of messenger ribonucleoprotein particles in somatic cells is a member of the Y-box binding transcription factor family. J. Biol. Chem., *270:* 3186–3192, 1995.

43. Matsumoto, K., Meric, F., and Wolffe, A. P. Translational repression dependent on the interaction of the *Xenopus* Y-box protein FRGY2 with mRNA. Role of the cold shock domain, tail domain, and selective RNA sequence recognition. J. Biol. Chem., *271:* 22706–22712, 1996.

44. Evdokimova, V., Ruzanov, P., Imataka, H., Raught, B., Svitkin, Y., Ovchinnikov, L. P., and Sonenberg, N. The major mRNA-associated protein YB-1 is a potent 5' cap-dependent mRNA stabilizer. EMBO J., *20:* 5491–5502, 2001.

45. Saito, H., Hayday, A. C., Wiman, K., Hayward, W. S., and Tonegawa, S. Activation of the c-*myc* gene by translocation: a model for translational control. Proc. Natl. Acad. Sci. USA, *80:* 7476–7480, 1983.

46. Nanbru, C., Lafon, I., Audigier, S., Gensac, M. C., Vagner, S., Huez, G., and Prats, A. C. Alternative translation of the proto-oncogene c-*myc* by an internal ribosome entry site. J. Biol. Chem., *272:* 32061–32066, 1997.

47. Stoneley, M., Paulin, F. E., Le Quesne, J. P., Chappell, S. A., and Willis, A. E. c-Myc 5' untranslated region contains an internal ribosome entry segment. Oncogene, *16:* 423–428, 1998.

48. Paulin, F. E., West, M. J., Sullivan, N. F., Whitney, R. L., Lyne, L., and Willis, A. E. Aberrant translational control of the c-*myc* gene in multiple myeloma. Oncogene, *13:* 505–513, 1996.

49. Chappell, S. A., LeQuesne, J. P., Paulin, F. E., de Schoolmeester, M. L., Stoneley, M., Soutar, R. L., Ralston, S. H., Helfrich, M. H., and Willis, A. E. A mutation in the c-myc-IRES leads to enhanced internal ribosome entry in multiple myeloma: a novel mechanism of oncogene de-regulation. Oncogene, *19:* 4437–4440, 2000.

50. Signori, E., Bagni, C., Papa, S., Primerano, B., Rinaldi, M., Amaldi, F., and Fazio, V. M. A. A somatic mutation in the 5'UTR of *BRCA1* gene in sporadic breast cancer causes down-modulation of translation efficiency. Oncogene, *20:* 4596–4600, 2001.

51. Liu, L., Dilworth, D., Gao, L., Monzon, J., Summers, A., Lassam, N., and Hogg, D. Mutation of the CDKN2A 5' UTR creates an aberrant initiation codon and predisposes to melanoma. Nat. Genet., *21:* 128–132, 1999.

52. Savitsky, K., Platzer, M., Uziel, T., Gilad, S., Sartiel, A., Rosenthal, A., Elroy-Stein, O., Shiloh, Y., and Rotman, G. Ataxia-telangiectasia: structural diversity of untranslated sequences suggests complex post-transcriptional regulation of *ATM* gene expression. Nucleic Acids Res., *25:* 1678–1684, 1997.

53. Brown, C. Y., Mize, G. J., Pineda, M., George, D. L., and Morris, D. R. Role of two upstream open reading frames in the translational control of oncogene *mdm2*. Oncogene, *18:* 5631–5637, 1999.

54. Sobczak, K., and Krzyzosiak, W. J. Structural determinants of BRCA1 translational regulation. J. Biol. Chem., *277:* 17349–17358, 2002.

55. Arrick, B. A., Grendell, R. L., and Griffin, L. A. Enhanced translational efficiency of a novel transforming growth factor β3 mRNA in human breast cancer cells. Mol. Cell. Biol., *14:* 619–628, 1994.

56. Myers, S. M., Eng, C., Ponder, B. A., and Mulligan, L. M. Characterization of *RET* proto-oncogene 3' splicing variants and polyadenylation sites: a novel C-terminus for RET. Oncogene, *11:* 2039–2045, 1995.

57. Byrne, J. A., Tomasetto, C., Rouyer, N., Bellocq, J. P., Rio, M. C., and Basset, P. The tissue inhibitor of *metalloproteinases-3* gene in breast carcinoma: identification of multiple polyadenylation sites and a stromal pattern of expression. Mol. Med., *1:* 418–427, 1995.

58. Moscow, J. A., He, R., Gudas, J. M., and Cowan, K. H. Utilization of multiple polyadenylation signals in the human *RHOA* protooncogene. Gene (Amst.), *144:* 229–236, 1994.

59. Senterre-Lesenfants, S., Alag, A. S., and Sobel, M. E. Multiple mRNA species are generated by alternate polyadenylation from the human *calmodulin-1* gene. J. Cell. Biochem., *58:* 445–454, 1995.

60. Lazaris-Karatzas, A., Montine, K. S., and Sonenberg, N. Malignant transformation by a eukaryotic initiation factor subunit that binds to mRNA 5' cap. Nature (Lond.), *345:* 544–547, 1990.

61. De Benedetti, A., and Rhoads, R. E. Overexpression of eukaryotic protein synthesis initiation factor 4E in HeLa cells results in aberrant growth and morphology. Proc. Natl. Acad. Sci. USA, *87:* 8212–8216, 1990.

62. Polunovsky, V. A., Rosenwald, I. B., Tan, A. T., White, J., Chiang, L., Sonenberg, N., and Bitterman, P. B. Translational control of programmed cell death: eukaryotic translation initiation factor 4E blocks apoptosis in growth-factor-restricted fibroblasts with physiologically expressed or deregulated Myc. Mol. Cell. Biol., *16:* 6573–6581, 1996.

63. Miyagi, Y., Sugiyama, A., Asai, A., Okazaki, T., Kuchino, Y., and Kerr, S. J. Elevated levels of eukaryotic translation initiation factor eIF-4E, mRNA in a broad spectrum of transformed cell lines. Cancer Lett., *91:* 247–252, 1995.

64. Kerekatte, V., Smiley, K., Hu, B., Smith, A., Gelder, F., and De Benedetti, A. The proto-oncogene/translation factor *eIF4E*: a survey of its expression in breast carcinomas. Int. J. Cancer, *64:* 27–31, 1995.

65. Li, B. D., Liu, L., Dawson, M., and De Benedetti, A. Overexpression of eukaryotic initiation factor 4E (eIF4E) in breast carcinoma. Cancer (Phila.), *79:* 2385–2390, 1997.

66. Sorrells, D. L., Black, D. R., Meschonat, C., Rhoads, R., De Benedetti, A., Gao, M., Williams, B. J., and Li, B. D. Detection of *eIF4E* gene amplification in breast cancer by competitive PCR. Ann. Surg. Oncol., *5:* 232–237, 1998.

67. Li, B. D., McDonald, J. C., Nassar, R., and De Benedetti, A. Clinical outcome in stage I to III breast carcinoma and eIF4E overexpression. Ann. Surg., *227:* 756–761; discussion, 761–763, 1998.

68. Wang, S., Rosenwald, I. B., Hutzler, M. J., Pihan, G. A., Savas, L., Chen, J. J., and Woda, B. A. Expression of the eukaryotic translation initiation factors 4E and 2α in non-Hodgkin's lymphomas. Am. J. Pathol., *155:* 247–255, 1999.

69. Fukuchi-Shimogori, T., Ishii, I., Kashiwagi, K., Mashiba, H., Ekimoto, H., and Igarashi, K. Malignant transformation by overproduction of translation initiation factor eIF4G. Cancer Res., *57:* 5041–5044, 1997.

70. Eberle, J., Krasagakis, K., and Orfanos, C. E. Translation initiation factor eIF-4A1 mRNA is consistently overexpressed in human melanoma cells *in vitro*. Int. J. Cancer, *71:* 396–401, 1997.

71. Shuda, M., Kondoh, N., Tanaka, K., Ryo, A., Wakatsuki, T., Hada, A., Goseki, N., Igari, T., Hatsuse, K., Aihara, T., Horiuchi, S., Shichita, M., Yamamoto, N., and Yamamoto, M. Enhanced expression of translation factor mRNAs in hepatocellular carcinoma. Anticancer Res., *20:* 2489–2494, 2000.

72. Nupponen, N. N., Porkka, K., Kakkola, L., Tanner, M., Persson, K., Borg, A., Isola, J., and Visakorpi, T. Amplification and overexpression of p40 subunit of eukaryotic translation initiation factor 3 in breast and prostate cancer. Am. J. Pathol., *154:* 1777–1783, 1999.

73. Rothe, M., Ko, Y., Albers, P., and Wernert, N. Eukaryotic initiation factor 3 p110 mRNA is overexpressed in testicular seminomas. Am. J. Pathol., *157:* 1597–1604, 2000.

74. Barlund, M., Forozan, F., Kononen, J., Bubendorf, L., Chen, Y., Bittner, M. L., Torhorst, J., Haas, P., Bucher, C., Sauter, G., Kallioniemi, O. P., and Kallioniemi, A. Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. J. Natl. Cancer Inst. (Bethesda), *92:* 1252–1259, 2000.

75. Topalian, S. L., Kaneko, S., Gonzales, M. I., Bond, G. L., Ward, Y., and Manley, J. L. Identification and functional characterization of neo-poly(A) polymerase, an RNA processing enzyme overexpressed in human tumors. Mol. Cell. Biol., *21:* 5614–5623, 2001.

76. Scorilas, A., Talieri, M., Ardavanis, A., Courtis, N., Dimitriadis, E., Yotis, J., Tsiapalis, C. M., and Trangas, T. Polyadenylate polymerase enzymatic activity in mammary tumor cytosols: a new independent prognostic marker in primary breast cancer. Cancer Res., *60:* 5427–5433, 2000.

77. Janz, M., Harbeck, N., Dettmar, P., Berger, U., Schmidt, A., Jurchott, K., Schmitt, M., and Royer, H. D. Y-box factor YB-1 predicts drug resistance and patient outcome in breast cancer independent of clinically relevant tumor biologic factors HER2, uPA and PAI-1. Int. J. Cancer, *97:* 278–282, 2002.

78. Shibahara, K., Sugio, K., Osaki, T., Uchiumi, T., Maehara, Y., Kohno, K., Yasumoto, K., Sugimachi, K., and Kuwano, M. Nuclear expression of the Y-box binding protein, YB-1, as a novel marker of disease progression in non-small cell lung cancer. Clin. Cancer Res., *7:* 3151–3155, 2001.

79. Kamura, T., Yahata, H., Amada, S., Ogawa, S., Sonoda, T., Kobayashi, H., Mitsumoto, M., Kohno, K., Kuwano, M., and Nakano, H. Is nuclear expression of Y box-binding protein-1 a new prognostic factor in ovarian serous adenocarcinoma? Cancer (Phila.), *85:* 2450–2454, 1999.

80. Bargou, R. C., Jurchott, K., Wagener, C., Bergmann, S., Metzner, S., Bommert, K., Mapara, M. Y., Winzer, K. J., Dietel, M., Dorken, B., and Royer, H. D. Nuclear localization and increased levels of transcription factor YB-1 in primary human breast cancers are associated with intrinsic *MDR1* gene expression. Nat. Med., *3:* 447–450, 1997.

81. Aoki, M., Blazek, E., and Vogt, P. K. A role of the kinase mTOR in cellular transformation induced by the oncoproteins P3k and Akt. Proc. Natl. Acad. Sci. USA, *98:* 136–141, 2001.

82. Child, S. J., Miller, M. K., and Geballe, A. P. Cell type-dependent and -independent control of HER-2/neu translation. Int. J. Biochem. Cell Biol., *31:* 201–213, 1999.

83. Jefferies, H. B., Reinhard, C., Kozma, S. C., and Thomas, G. Rapamycin selectively represses translation of the "polypyrimidine tract" mRNA family. Proc. Natl. Acad. Sci. USA, *91:* 4441–4445, 1994.

84. Terada, N., Patel, H. R., Takase, K., Kohno, K., Nairn, A. C., and Gelfand, E. W. Rapamycin selectively inhibits translation of mRNAs encoding elongation factors and ribosomal proteins. Proc. Natl. Acad. Sci. USA, *91:* 11477–11481, 1994.

85. Jefferies, H. B., Fumagalli, S., Dennis, P. B., Reinhard, C., Pearson, R. B., and Thomas, G. Rapamycin suppresses 5'TOP mRNA translation through inhibition of p70s6k. EMBO J., *16:* 3693–3704, 1997.

86. Beretta, L., Gingras, A. C., Svitkin, Y. V., Hall, M. N., and Sonenberg, N. Rapamycin blocks the phosphorylation of 4E-BP1 and inhibits cap-dependent initiation of translation. EMBO J., *15:* 658–664, 1996.

87. Hidalgo, M., and Rowinsky, E. K. The rapamycin-sensitive signal transduction pathway as a target for cancer therapy. Oncogene, *19:* 6680–6686, 2000.

88. Hosoi, H., Dilling, M. B., Shikata, T., Liu, L. N., Shu, L., Ashmun, R. A., Germain, G. S., Abraham, R. T., and Houghton, P. J. Rapamycin causes poorly reversible inhibition of mTOR and induces p53-independent apoptosis in human rhabdomyosarcoma cells. Cancer Res., *59:* 886–894, 1999.

89. Huang, S., and Houghton, P. J. Resistance to rapamycin: a novel anticancer drug. Cancer Metastasis Rev., *20:* 69–78, 2001.

90. Geoerger, B., Kerr, K., Tang, C. B., Fung, K. M., Powell, B., Sutton, L. N., Phillips, P. C., and Janss, A. J. Antitumor activity of the rapamycin analog CCI-779 in human primitive neuroectodermal tumor/medulloblastoma models as single agent and in combination chemotherapy. Cancer Res., *61:* 1527–1532, 2001.

91. Gibbons, J. J., Discafani, C., Peterson, R., Hernandez, R., Skotnicki, J., and Frost, P. The effect of CCI-779, a novel macrolide anti-tumor agent, on the growth of human tumor cells *in vitro* and in nude mouse xenografts *in vivo*. Proc. Am. Assoc. Cancer Res., *40:* 301, 1999.

92. Hidalgo, M., Rowinsky, E., Erlichman, C., Marshall, B., Marks, R., Edwards, T., and Buckner, J. J. A Phase I and pharmacological study of CCI-779 cycle inhibitor. Ann. Oncol., *11* (Suppl. 4): 133, 2001.

93. Alexandre, J., Raymond, E., Depenbrock, H., Mekhaldi, S., Angevin, E., Paillet, C., Hanauske, A., Frisch, J., Feussner, A., and Armand, J. P. CCI-779, a new rapamycin analog, has antitumor activity at doses inducing only mild cutaneous effects and mucositis: early results of an ongoing Phase I study. Proceedings of the 1999 AACR-NCI-EORTC International Conference, Clin. Cancer Res., *5* (Suppl.): 3730s, 1999.

94. Chan, S., Johnston, S., Scheulen, M. E., Mross, K., Morant, A., Lahr, A., Feussner, A., Berger, M., and Kirsch, T. First report: a Phase 2 study of the safety and activity of CCI-779 for patients with locally advanced or metastatic breast cancer failing prior chemotherapy. Proc. Am. Soc. Clin. Oncol., *21:* 44a, 2002.

95. Atkins, M. B., Hidalgo, M., Stadler, W., Logan, T., Dutcher, J. P., Hudes, G., Park, Y., Marshall, B., Boni, J., and Dukart, G. A randomized double-blind Phase 2 study of intravenous CCI-779 administered weekly to patients with advanced renal cell carcinoma. Proc. Am. Soc. Clin. Oncol., *21:* 10a, 2002.

96. Smith, S. G., Trinh, C. M., Inge, L. J., Thomas, G., Cloughsey, T. F., Sawyers, C. L., and Mischel, P. S. PTEN expression status predicts glioblastoma cell sensitivity to CCI-779. Proc. Am. Assoc. Cancer Res., *43:* 335, 2002.

97. Yu, K., Toral-Barza, L., Discafani, C., Zhang, W. G., Skotnicki, J., Frost, P., and Gibbons, J. J. mTOR, a novel target in breast cancer: the effect of CCI-779, an mTOR inhibitor, in preclinical models of breast cancer. Endocr. Relat. Cancer, *8:* 249–258, 2001.

98. Dilling, M. B., Germain, G. S., Dudkin, L., Jayaraman, A. L., Zhang, X., Harwood, F. C., and Houghton, P. J. 4E-binding proteins, the suppressors of eukaryotic initiation factor 4E, are downregulated in cells with acquired or intrinsic resistance to rapamycin. J. Biol. Chem., *277:* 13907–13917, 2002.

99. Guba, M., von Breitenbuch, P., Steinbauer, M., Koehl, G., Flegel, S., Hornung, M., Bruns, C. J., Zuelke, C., Farkas, S., Anthuber, M., Jauch, K. W., and Geissler, E. K. Rapamycin inhibits primary and metastatic tumor growth by antiangiogenesis: involvement of vascular endothelial growth factor. Nat. Med., *8:* 128–135, 2002.

100. Lane, H. A., Schell, C., Theuer, A., O'Reilly, T., and Wood, J. Anti-angiogenic activity of RAD001, an orally active anticancer agent. Proc. Am. Assoc. Cancer Res., *43:* 184, 2002.

101. Maeshima, Y., Sudhakar, A., Lively, J. C., Ueki, K., Kharbanda, S., Kahn, C. R., Sonenberg, N., Hynes, R. O., and Kalluri, R. Tumstatin, an endothelial cell-specific inhibitor of protein synthesis. Science (Wash. DC), *295:* 140–143, 2002.

102. Caygill, C. P., Charlett, A., and Hill, M. J. Fat, fish, fish oil and cancer. Br. J. Cancer, *74:* 159–164, 1996.

103. Falconer, J. S., Ross, J. A., Fearon, K. C., Hawkins, R. A., O'Riordain, M. G., and Carter, D. C. Effect of eicosapentaenoic acid and other fatty acids on the growth *in vitro* of human pancreatic cancer cell lines. Br. J. Cancer, *69:* 826–832, 1994.

104. Noguchi, M., Minami, M., Yagasaki, R., Kinoshita, K., Earashi, M., Kitagawa, H., Taniya, T., and Miyazaki, I. Chemoprevention of DMBA-induced mammary carcinogenesis in rats by low-dose EPA and DHA. Br. J. Cancer, *75:* 348–353, 1997.

105. Palakurthi, S. S., Fluckiger, R., Aktas, H., Changolkar, A. K., Shahsafaei, A., Harnett, S., Kilic, E., and Halperin, J. A. Inhibition of translation initiation mediates the anticancer effect of the n-3 polyunsaturated fatty acid eicosapentaenoic acid. Cancer Res., *60:* 2919–2925, 2000.

106. Aktas, H., Fluckiger, R., Acosta, J. A., Savage, J. M., Palakurthi, S. S., and Halperin, J. A. Depletion of intracellular Ca^{2+} stores, phosphorylation of eIF2α, and sustained inhibition of translation initiation mediate the anticancer effects of clotrimazole. Proc. Natl. Acad. Sci. USA, *95:* 8280–8285, 1998.

107. Mhashilkar, A. M., Schrock, R. D., Hindi, M., Liao, J., Sieger, K., Kourouma, F., Zou-Yang, X. H., Onishi, E., Takh, O., Vedvick, T. S., Fanger, G., Stewart, L., Watson, G. J., Snary, D., Fisher, P. B., Saeki, T., Roth, J. A., Ramesh, R., and Chada, S. Melanoma differentiation associated gene-7 (*mda-7*): a novel anti-tumor gene for cancer gene therapy. Mol. Med., *7:* 271–282, 2001.

108. Su, Z. Z., Madireddi, M. T., Lin, J. J., Young, C. S., Kitada, S., Reed, J. C., Goldstein, N. I., and Fisher, P. B. The cancer growth suppressor gene *mda-7* selectively induces apoptosis in human breast cancer cells and inhibits tumor growth in nude mice. Proc. Natl. Acad. Sci. USA, *95:* 14400–14405, 1998.

109. Saeki, T., Mhashilkar, A., Chada, S., Branch, C., Roth, J. A., and Ramesh, R. Tumor-suppressive effects by adenovirus-mediated *mda-7* gene transfer in non-small cell lung cancer cell *in vitro*. Gene Ther., *7:* 2051–2057, 2000.

110. Pataer, A., Vorburger, S. A., Barber, G. N., Chada, S., Mhashilkar, A. M., Zou-Yang, H., Stewart, A. L., Balachandran, S., Roth, J. A., Hunt, K. K., and Swisher, S. G. Adenoviral transfer of the melanoma differentiation-associated gene 7 (*mda7*) induces apoptosis of lung cancer cells via up-regulation of the double-stranded RNA-dependent protein kinase (PKR). Cancer Res., *62:* 2239–2243, 2002.

111. Ito, T., Warnken, S. P., and May, W. S. Protein synthesis inhibition by flavonoids: roles of eukaryotic initiation factor 2α kinases. Biochem. Biophys. Res. Commun., *265:* 589–594, 1999.

112. Eberle, J., Fecker, L. F., Bittner, J. U., Orfanos, C. E., and Geilen, C. C. Decreased proliferation of human melanoma cell lines caused by antisense RNA against translation factor eIF-4A1. Br. J. Cancer, *86:* 1957–1962, 2002.

113. Polunovsky, V. A., Gingras, A. C., Sonenberg, N., Peterson, M., Tan, A., Rubins, J. B., Manivel, J. C., and Bitterman, P. B. Translational control of the antiapoptotic function of Ras. J. Biol. Chem., *275:* 24776–24780, 2000.

114. D'Cunha, J., Kratzke, M. G., Alter, M. D., Polunovsky, V. A., Bitterman, P. B., and Kratzke, R. A. Over-expression of the translational repressor 4E-BP1 inhibits NSCLC tumorigenicity *in vivo*. Proc. Am. Assoc. Cancer Res., *43:* 816–817, 2002.

115. DeFatta, R. J., Nathan, C. A., and De Benedetti, A. Antisense RNA to eIF4E suppresses oncogenic properties of a head and neck squamous cell carcinoma cell line. Laryngoscope, *110:* 928–933, 2000.

116. Herbert, T. P., Fahraeus, R., Prescott, A., Lane, D. P., and Proud, C. G. Rapid induction of apoptosis mediated by peptides that bind initiation factor eIF4E. Curr. Biol., *10:* 793–796, 2000.

117. DeFatta, R. J., Li, Y., and De Benedetti, A. Selective killing of cancer cells based on translational control of a suicide gene. Cancer Gene Ther., *9:* 573–578, 2002.